

Quality in the gathering of online data: Standardising online questionnaires, integration with administrative sources and development of bias control mechanisms (EUSTAT)

Name(s) of the author (s): Cristina Prado¹, Patxi Pizarro², Carmen Guinea³

¹ *Basque Statistics Office - Eustat, Vitoria, Spain; Cristina_prado@eustat.eus*

² *Basque Statistics Office - Eustat, Vitoria, Spain; p-pizarro@eustat.eus*

³ *Basque Statistics Office - Eustat, Vitoria, Spain; mcarmen_guinea@eustat.eus*

Abstract

Eustat has spent years gathering data for statistical purposes over the Internet and carrying out a number of projects related to the standardisation of web questionnaires, their integration with data from administrative sources and the development of processes that improve the quality of the data gathered and reduce non-response bias. This document will explain the results of these projects and their practical applications in the data gathering process.

Keywords: Quality, design, integration, bias, control.

1. Introduction

The Basque Statistics Institute has spent years gathering data over the Internet, and has managed to ensure that all surveys, whether economic, demographic or social, now use this collection method, which has led not only to a reduction in costs, but also an improvement in the management of human resources dedicated to these tasks. It also gives all citizens a quick, simple and convenient way of completing surveys.

Furthermore Eustat, aware of the huge importance of data quality in statistical processes, has made a great effort to standardise the design of all web questionnaires, to improve their accessibility, create designs that are more visual, and offer help and controls to avoid the biases that may arise in data from self-completed surveys, which do not have the external help present in other collection methods.

Another course of action undertaken in recent years has been to integrate data from external administrative sources into web questionnaires, to serve a dual function. Firstly, it serves to contrast and verify the data entered in the questionnaires, and secondly, it is used to give respondents suggestions for possible responses. This allows us to improve the quality of the responses and reduce non-responses for certain types of questions.

Finally, we also want to show the Demographic Survey Manager that we have developed a single application that controls and monitors all of the surveys from a single point, and allows us to obtain data on the status of the survey process online, allowing us to establish corrective measures against unexpected situations that affect the quality of the data being collected at that moment.

The aim of the report will be to demonstrate the standardised design of Eustat's web questionnaires from the perspective of data quality, and not only regarding the format of the questions but also regarding the control and error validation system, the help system established to facilitate question comprehension, and the integration of data from administrative sources from external bodies. To do this, we will use various online questionnaires that are currently live, in the areas of economics and demographics, namely

"Private Education Expenditure and Financing", "The Information Society for Families" and "Population in Relation to Activity".

2. Standardisation of web questionnaires

The main objective of Eustat's internet standardisation project was to create a repository of software components that would allow us to develop web questionnaires within a single basic framework for all the pages that make up a survey, including the following features: Organisation and hierarchy of pages, common style sheets for all the surveys to ensure each questionnaire has a standard design, basic browsing and menus, a unique identification system and access control, a general system for the interaction of feedback messages, validations, error and warning control and the standardisation of different data viewing modes in order to make completing the questionnaires simpler. All these features have been developed whilst bearing in mind the different issues that might have an impact on the quality of the data, especially:

- **Usability:** The questionnaire must be user-friendly and easy to handle so as to avoid errors and difficulties when it comes to entering data, and in order to ensure respondents do not use it incorrectly. Furthermore, it must also be sufficiently intuitive so that it does not matter how good the ICT skills of the respondents are.
- **Self-management:** The respondents do not have any outside help to guide them and deal with any questions they might have when they are completing the questionnaire. As such, the level of clarity of the questions and answers as well as the help system must strike a balance on an ICT and data validation level alike between exhaustiveness and design efficiency. This is to ensure that the data is high quality, but not at the cost of redundant data that has the opposite effect.

2.1. General design of the web questionnaires

It goes without saying that the design of the internet questionnaires is crucial in terms of the quality of the data being gathered. This is primarily because it is the respondents themselves

who handle it, which means they rely exclusively on the information and processes the questionnaires contain. Hence Eustat has designed a range of different features to ensure the questions and answers are understood with the clear goal of avoiding any potential confusion that would result in poor-quality responses, including non-responses, and optimising the system's general usability so that respondents do not give up in the middle of the questionnaire. All Eustat's web questionnaires have the following characteristics:

- a) A design based on the Eustat Style Guide, which sets out all the design and accessibility guidelines all questionnaires have to follow so that they contain common, consistent features. This applies to image design, typography, questionnaire browsing, page size, download speed, supported browsers and resolutions, generic page definitions, headers and so on.
- b) Logging in via the single entry page, including a user identification system, general information, specific information on each survey and features that let you change your password in order to increase log-in security:

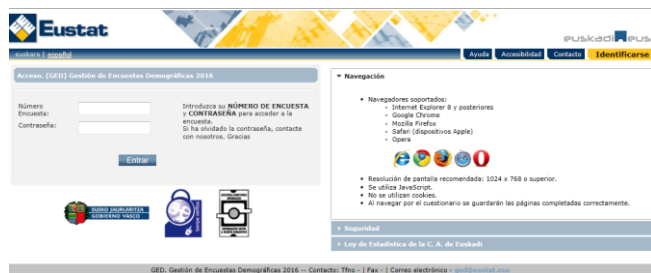


Fig. 1

The role of this page is to increase security and the confidence the respondents have in the confidentiality of their data.

- c) Definition of a homepage before the questionnaire begins. This page contains general information, manuals and explanatory videos that Eustat provides its users with to help them fill in the questionnaire.



Fig. 2

The purpose of this page is to explain to users what the institute does so that they are better equipped to fill in the forms and so that they are made aware of the importance of their responses to society.

- d) The layout of the pages has a single design to ensure the questionnaire is consistent across the board: Headers, footers, the arrangement of the different questionnaire tabs, the information and help section and the section for communicating with the respondent for suggestions and correcting errors.

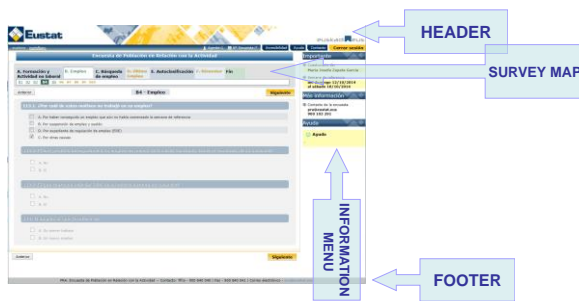


Fig. 3

The aim is to have a single framework for all the pages of the surveys by coming up with a layout and design that guarantees clarity and user-friendliness with room for constant improvement.

- e) Methods used to define the responses using the most suitable visual elements for any question type in order to make filling in questionnaires faster and to reduce the number of incorrect responses.



Fig 5.1 Using exclusive responses, enabling and disabling questions/responses according to previous responses in order to limit responses to correct ones only.



Fig 5.2 Mandatory questions ensure users cannot skip to the next page, which means key questions do not go unanswered.



Fig 5.3 Radio buttons are used for scale-type questions so that having the responses displayed helps users to complete the questionnaire.

The design of the questions/responses took into account recommendations by Mick P. Couper (2011) because of his research work and the conclusions he reached in the project ESSnet DCSS (Data Collection in Social Surveys Using Multiple Modes), a project he completed within Eurostat.

f) The data entered by the respondent is secured using printing mechanisms, and is saved as a PDF from the completed survey. This guarantees the security of the user's responses.



Fig. 5

g) Defining a graph control system that monitors any pages that are skipped, activated or deactivated during a survey. These may occur in relation to question responses, specific questions, question blocks, sections on the same page or entire pages, the activation/deactivation of which is subject to the response given to one or some of the previous questions.



Fig. 6

This function ensures that respondents only answer the questions we are interested in wherever possible. This helps us prevent respondents from giving up or giving incorrect answers as a result of the excessive length of the questionnaire.

2.2. Validation and error control system

Different levels of error control and validation have been defined in order to guarantee the best possible answers. Firstly, there are page validations, which are validations of related question responses on the same page. Secondly, there are cohesion validations, which are validations of questions from different pages so as to prevent the questionnaire from being completed if there are serious errors in this vein, thereby ensuring the responses are complete. Then there are length validations, which are validations for information purposes that are based on knowledge Eustat has, and which will not have an effect on the questionnaire's completion. Additionally, each error can be “soft” or “hard”, the idea being that the respondent is simply “forced” to correct important data we know is not correct.

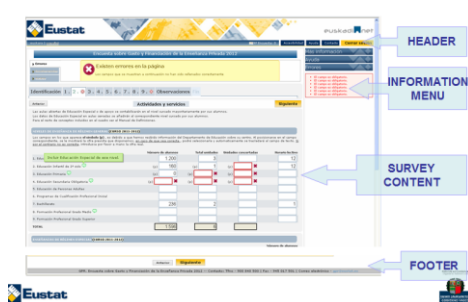


Fig. 7

The message viewing system has been set up in a very visual way, and it differs according to the significance of the error so that respondents can be more intuitive when it comes to making the corrections, and to provide the respondent with all the information we have. To add to this, particular emphasis has been placed on how often error messages appear and the process of finding them. For each instance, the most appropriate options have been defined. Generally

speaking, the error messages can be viewed as personal messages in the field the user is in or on the information panel so as to provide the respondent with both an individual and overall view of what they have to correct.

2.3. Information and help system

The information respondents are provided with in the forms is crucial in ensuring there are no misinterpretations that may affect the quality of the responses. Indeed, it is these forms that take the place of the survey taker in terms of helping respondents to complete the form. As such, the information must be very visible and easy to access. To this end, the system defined includes different levels of explanations so as to provide the respondent with constant aid so that they reach the objectives set for each question. The first level is defined in the text of the question being answered, and includes broader explanations that appear in pop-up windows. The second level is defined by each response using the same system the first level uses. The third level consists of information explained in the control panel, and it may contain further documentation that is related to the questions on each page in addition to explanatory videos. Depending on the relevance of the information, either one level, two levels or all three levels are used.

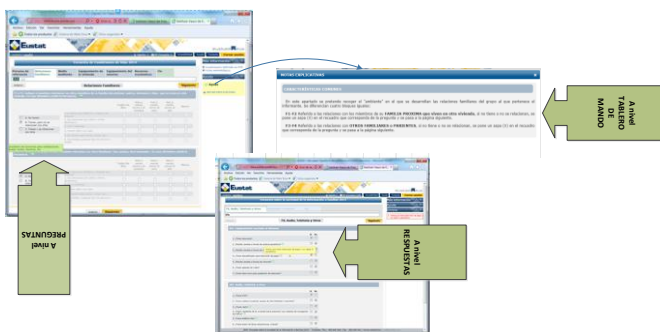


Fig. 8

The aim of this system is to ensure the respondents always have access to information that will clear up any issues they may have during the process.

3. Integration of administrative sources

Nowadays, all information is managed using ICT resources, meaning there are numerous administrative sources. Eustat is aware of how important the internet is, and has integrated such data into web questionnaires with a view not only to improve the quality and reliability of the data gathered, but also to streamline and reduce data collection costs. Hence, once the data contained in the sources - data that in fact undergoes a conversion from administrative sources into the statistical sources - has been validated, filtered, identified and homogenised, we then use these sources (following Daniel Lewis' recommendations) not only to conduct surveys based exclusively on administrative data, but we also use them as basic complementary data within the web surveys in the following way:

- a) Shortening surveys by omitting questions.
- b) Suggested responses: When we have highly reliable data, but we need a confirmation from the respondent, we include this data in the relevant web questionnaire, and we have the option of selecting or modifying the data. An example of this feature is the survey "Private Education Expenditure and Financing". In this survey, we use data provided by the Basque Government's Department of Education:



Fig. 9

- c) Validation of responses: Using the data to increase the reliability of responses by making "hot" contrasts between the response given and the one Eustat has. Data originating from different sources (Social Security and the Basque Employment Institute) was integrated into the "Labour Force" survey:



Fig. 10

4. Application for controlling the survey process

Another of the processes that influence the quality of the data are tasks carried out during the survey period that are directly linked to the professionalism of the people involved in the process and the resources available. To standardise all the tasks that the different survey companies carry out, a web application has been developed that will act as an entry point for all survey takers, where they will be provided with the ICT processes necessary to complete their work. This includes “hot” validation processes that will allow them to detect potential defects in the quality of the data or low response rates so that they are able to come up with the corrective measures required straight away:

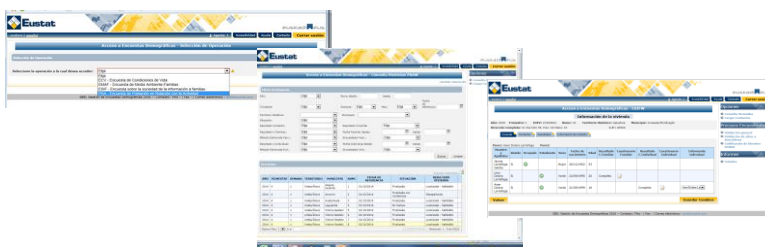


Fig. 11

Furthermore, in the case of demographic surveys, questions about housing components have been rephrased with a view to select the reference person by making use of the new technology and by replacing the survey format with a more user-friendly format that will ensure errors are not made when completing it:



Fig. 12

In other words, we take full advantage of the technology, and we transform the concept of the conventional survey into a type of survey that is more versatile and, of course, one that respondents complete independently.

5. Satisfaction Survey

To obtain constant feedback from the respondents, as well as information about areas where the questionnaires and the processes in general could improve, the system allows you to add questions to satisfaction surveys that you are interested in at that moment for a certain questionnaire. The format is as follows:

La encuesta se ha dado por finalizada

	Muy mala	Mala	Regular	Buena	Muy mala
Pregunta 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
Pregunta 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pregunta 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

This survey allows us to make continuous improvements with a view to optimise the web questionnaires and strengthen ties with respondents. We believe this in turn results in better quality data, and it is also improves the general image of Statistical Institutes.

6. References

Web Survey Methodology: Interface Design, Sampling and Statistical Inference. Mick P. Couper, 2011. Research professor at the Survey Research Centre.

ESSnet DCSS – Data Collection in Social Surveys Using multiple modes, Eurostat 2014-2015

The use of Administrative and Accounts Data for Economic Statistics, Daniel Lewis, 2014