UNITED NATIONS ECONOMIC
COMMISSION FOR EUROPE (UNECE)

CONFERENCE OF EUROPEAN
STATISTICIANS

EUROPEAN COMMISSION

STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

NEW ONLINE SERVICES FOR ACCESSING PHC MICRODATA FOR ANY USER

Lucian Alexandrescu

lucian.alexandrescu@insse.ro

National Institute of Statistics, Romania

Abstract:

Due to the multitude and diversity of census data requirements, NSI Romania (National Statistics Institute) developed a Query System Application (QSA) for PHC (Population and Household Census) data through which it is offered free online access to any user.

The QSA has a very friendly interface, the users being guided, step by step, to select the census year, the domain (population, dwellings or households), the geographical level, area, gender, age, educational level etc, till the statistical indicators, as these were registered in the census forms by the enumerator.

PHC database contains now only data from 1992 and 2002 Censuses and, as soon as possible, will be updated with 2011 PHC results.

The QSA offers access to a sample of 10% from the total amount of PHC 1992 and 2002 microdata. Any user can define queries, build and tailor his own statistical tables online through this QSA.

This application has facilities for saving the user's queries that can be run on the whole amount of data by NSI personnel and has multiple export formats.

1.   Introduction

Population and Household Census (PHC) represents the most complex and extensive statistical research that provides the essential information underlying justification of all projects and programs on macro and micro economic scale for at least a decade.

The multitude of information, the diversity of indicators, detailing them to the smallest territorial administrative unit, are all arguments for NSI to have been motivated to have a

Query System Application (QSA) capable of allowing a quick and efficient query of the huge database resulting from Censuses.

The solution for this imperative was achieved by designing this QSA. The application was designed based on the principle that a statistical classification is a strict and detailed hierarchy which makes it possible to present information from different levels of aggregation depending on the validity date with full coverage on the observed field. Comply with this requirement, each analyzed phenomenon or object is placed into a single entity type classification, the analyzed objects in different branches are governed by uniform methodological principles.

The Query System Application for PHC is used by NSI personnel but also by public users. The NSI personnel have an „internal" module used to disseminate the Censuses rezults and to solve the public users requests. Public users have access to a sample of 10% of total amount of PHC microdata and all aggregated data available on NSI website.

Synthesizing, this application identifies four types of users:

- *Administrator* – is responsible with the administration of the application, having access to all its functionality.
- *Advanced user* – is an internal user, from NSI personel, which can create queries where smallest aggregate level possible is at village level.
- *Normal user* – is an internal user, from NSI personel, which can create queries where smallest aggregate level possible is at commune level.
- *Public user* – is an external user, who can access the public module of the application, available on NSI website. In order to have access to the functionality of the public module of the application it is necesary to obtain an username and a password. He can create his own queries achieving a narrow set of data (10%). He also can obtain the already aggregated data available to the public.

2. PHC application Structure

The Query System Application for PHC can be accessed by typing in the address bar of your browser the following web address: http://colectaredate.insse.ro/phc/public.do?siteLang=en .

2.1 Presentation page (Home page)

When accessing QSA the presentation page is displayed containing information regarding this application. On the top bar are three buttons: *access aggregated data, contact* and *help*.

Access aggregated data – allows the user to access the already aggregated data.

Contact – offers the posibility to contact the NSI personel.

Help – offers information about user instructions.

On this page are three options available. The first one – ***log in*** – for the users who already have a valid account. The second – ***register*** – for the new users and the third option – ***recover password*** – for the users who have forgotten their password.

2.2   Log in page

This page is displayed when accessing hyperlink „***Log in***" from home page. On this page the user will fill the fields „username" and „password" and then he will click the „OK" button. If in one of the two fields are entered false data, it will be displayed an error message whereby the user is informed that the login can not be performed. In case user wishes to cancel the login, he should click the "Cancel" button. After a successfull login, the **<u>Main Page</u>** will be displayed.

2.3   Registration page

This page is displayed when the hyperlink "Register" from the **<u>Home page</u>** is clicked by any user wishing to obtain a valid account for the PHC application.

For obtaining a valid account, the user will have to fill in the mandatory informations: first Name, Last Name, E-mail, Password, Re-type Password, Adress, City, Country and Verification Code. The user can also fill additional information for a better communication, that are: County, Phone number, Fax, Company name, Company address and Function within the company.

In case the user fails to fill one or more mandatory information or inserts invalid data, the registration cannot be continued and the application shows an appropiate message and signals error reasons.

Only by correcting all errors, the account registration can be continued.
After filling mandatory data, the user will continue by clicking the "OK" button in order to regster the account.

The system will create a new account and will send a confirmation e-mail to the supplied e-mail address, containing all required information to activate the newly created account. This is neccessary in order to prevent abusses or registration made by automated systems.

In case one wishes the cancellation of the registration, the button "Cancel" should be clicked.

2.4   Main page

The page is displayed only after a successfull login.

The page displays information regarding the data that can be obtained using the PHC application. Also, the main menu of the application is activated in this page and exposes the following menu items: Create Query, Modify Query, Delete Query, Execute Query, Access Aggregated Data, Metada, Contact, Help, Log out, User information.

2.4.1. Create Query

The page can be accesssed only by logged users, by accessing "Create Query" from the menu. Creating a query involves several steps, the users being guided by the wizard. The first step of the wizard is ***selecting a query subject***. Fields mark with (*) are mandatory.

a)   Select subject.

In order to get to the next step in creating a query, the user must select a census year and a subject and press the "Next" button.

In case the user does not select any subject, the application will display an error message that a subject must be chosen in order to continue.

In case the user wants to cancel the query creation, the user should press the "Cancel" button.

b)   Select aggregation criteria

After selecting the subject at the previous step, in order to continue the wizard, the user must select the ***aggregation criteria*** available for the selected subject.

It is mandatory that the user selects at least one criteria for each table dimension (rows and columns).

The user will select the desired criterion, and by using the "Column >>" or "Row >>" buttons the item will be moved to teh column or row of the table. In order to deselect a criterion, the user will select the item and the one of the buttons "<< Column" and "<< Row" respectively will remove the citerion from the column or row and will be available in the criteria list.

To order the aggregation criteria, the user will select the criterion and will press one of the "Up" or "Down" buttons, depending on the user's wish.

In case the user wishes to obtain the total on rows or columns, on each of the table dimentions, he/she will select the "Total" checkbox available near to the columns section as well as the rows section.

After selecting and ordering of the criteria, in order to proceed to the next step, the user will click the "Next" button.

If the user wants to return to the previous screen, he should click the "Back" button.
In case he wishes to cancel the operation of creating a query, the user should press the "Close" button.

If the user presses the "Next" button without selecting any aggregation criteria the system will display a warning message to indicate that criteria must be selected, on both columns and rows.

c) Select aggregation values

After selecting the aggregation criteria, in order to continue through the wizard the user should select the available aggregation values for the already selected criteria.
It is mandatory that at least a value is selected for each criterion.
Multiple values for an aggregation criterion may be selected by clicking and holding the "CTRL" key at the same time.

After selecting aggregation values, in order to continue through the next step, the user should press the "Next" button.

If the user wishes to return to the previuos step he should click the "Back" button.
In case he wishes to cancel the operation of creating a query, the user should press the "Close" button.

If the user presses the "Next" button without selecting an aggregation value, the system shows a warning message to indicate that aggregation values should be selected.

d) Select statistical indicators

After selecting the aggregation values, in order to continue through the wizard the user should select the statistical indicators for which the query is build.
It is mandatory that the user select one of the two options:

- Number of objects
- Other indicators

After selecting one of the options, the user will press the "Next" button to continue through the wizard.

If the user wishes to return to the previuos step he should click the "Back" button.
In case he wishes to cancel the operation of creating a query, the user should press the "Close" button.

If the user presses the "Next" button without selecting any of the two options, the system will display a warning message.

e) Select indicators range

In case during the last step the "Number of objects" option was selected, an additional step is displayed. Here, the user sets the microdata filtering criteria for which the aggregation of the query will be performed.

Available functions to define the conditons: ABS, ROUND şi TRUNC.
Available operators to define the conditions are: =, >, >=, <, <=, <>, LIKE, IN, BETWEEN.
After selecting the column, the function the operator and the desired values, the user should press the „Add" button, and the defined condition will be displayed in the list.

To delete a condition from the list, the "Delete" checkbox available on the conditions side should be checked and then the "Delete" button clicked by the user.

The "Value" field cand be filled in by either writing the value directly or by selectin a value from a list of possible values. The list of values can be obtained by pressing "Values List..." button . A window where the selection is made will be displayed in which the user chooses the desired values by checking the associated checkbox. The content of the list depends on the selected column.

After selecting the desired values the "Close" button should be closed, and the system automatically fils the selected values in the "Value" fields. By pressing the "Add" button, the new condition will be available in the conditions list.

Ater completing the field in the view with the corresponding values, the user should press the "Next" button to continue through the wizard.

If the user wishes to return to the previuos step he should click the "Back" button.
In case he wishes to cancel the operation of creating a query, the user should press the "Close" button.

f) Select other indicators

In case at the previous step the option "Other indicators" is checked, an additional step is displayed to the user where he selects the indicators list to the included in the query (available in the list of field "Column") and the aggregated function that will be applied on it. Also, an Alias for that indicator can be defined.

Available aggregated functions are: SUM, MAX, MIN, AVG and COUNT.
After selecting the indicator, aggregated function and the alias, the "Add" button should be clicked, and the indicator will be displayed in the list.

To delete an indicator from the list, the "Delete" corresponding chekbox should be checked and the "Delete" button pressed.

After completing the values from the view with the ones corresponding to the user's desires, he should press the "Next" button in oder to continue through the wizard.

g) Define results table title

Here the user defines the title and the region of the table that contains the result of the query.

If the user does not fill theese fields but still presses the next button, the system displays a warning message.

After filling the values for the two field within the view, the user should press the "Next" button to go to the last step of the wizard, "Save query request".

h) Save query

In this view the user sees a brief pressentation of the selection made and offers the possibility of executing the query. The result of the query can be displayed directly on the screen or saved in XLS, XML or CSV format.

Before executing the query, the user must fill in the values for the next fields: Query Name, Category and Access. The Category and Access fields are filled implicitely by the system with the first available fields in their associated lists. In case the user desires some other values, he will select one of the ones available in the selection list.

If the Private value is selected for the Access field, the query will be available only to the user creating it. If the Shared value for the Access field is selected, the query will be available to all users.

The Query name field must be filled by the user. If not supplied, if any of the available buttons in the screen in pressed , the system shows a warning message to inform the user that the query name should be supplied.

If the user wishes to go back to the previous screen, the "Back" button should be pressed.

If the "Execute" button is pressed the system executes the query and the result will be displayed in a new window.

If the "Export Excel" button is pressed, the system will export the result of the query into an Excel file.

If the "Export XML" button is pressed, the system will export the result of the query into a XML file.

If the "Export CSV" button is pressed, the system will export the result of the query into a CSV file.

If the "Save" button is pressed the system will save the query in the database and will display to the user a successfull message.

2.4.2 Modify query

This option offers the user the possibility of editing an own query.

If there are queries that match the search criteria chosen, the system will present them as a list.

This option is similar to the query creation opperation, the steps being the same as the ones involved in the query creation. This time the default fields are filled with the existing values in the selected query, the user being able to modify them.

2.4.3. Delete query

This option offers the user the possibility of deleting an own query.

The user will perform a search using some searching criteria.
To search a query, the user can use along the "Census" criteria only one of the following

criteria: "Category", "Name", "Date Last Modified","Last Execution Date" i.e. the options have been associated a radio button in the view.

After selecting the year of the census, the user selects one of the remaining criteria, fills a value for the selected criteria and clicks the "Search" button. If no query is found, the system displays "No results have been found".

In case the are queries that correspond to the selected criteria, the system will display them in a list.

To delete a query, the user will select one from the previously obtained list by selecting the associated radio button and clicking the "Delete" button. The system will display a confirmation message to the user to complete the opperation.

After the confirmation is given, the system deletes the selected query, updates the content of the view "Delete Query" and displays a success message. If the user wishes to cancel the deletion of a query, he should click the "Close" button.

2.4.4. Execute query

The options offers the user the possibility to execute queries.

The user will perform a search using some searching criteria.

To search a query, the user can use along the "Census" criteria only one of the following criteria: "Category", "Name", "Date Last Modified","Last Execution Date" i.e. the options have been associated a radio button in the view.

After selecting the year of the census, the user selects one of the remaining criteria, fills a value for the selected criteria and clicks the "Search" button. If no query is found, the system displays "No results have been fiund".

In case the are queries that correspond to the selected criteria, the system will display them in a list.

To execute a query, the user will select one from the previously obtained list by selecting the associated radio button and should click the "Execute" button. The system will run the query and display the results in a new screen.

If the user wishes to cancel the execution of a query, he should click the "Close" button.

2.4.5. Access aggregated data

In order to obtain aggregated data, the user must select category and result table and then click on „ view" button. The data table is save in .xls format and the file can be found in downloads folder (if it is used the mozilla firefox or google chrome, if it is used the IE browser, appear the file download window and the user can choose between „open", „save" or „cancel" buttons).

2.4.6. Metadata

In this page is presented the documentation for 1992 and 2002 Censuses. There is also a hyperlink from where microdata can be accessed (http://www.iecm-project.org/).

3.   Conclusions

As I hope it could be seen from the above, in the permanently effort to make the essential statistical information reach to the users as quickly and detailed as possible, National

Statistics Institute has now an effective tool:PHC application. With this tool any user can „scan" the richest deep statistical database available in any country, because no other statistical research than Census it does not lead to a large set of indicators, for the entire national area and fully comparable by simultaneity collecting. Certainly, the operating and perfectioning process of PHC is far from being over. He still supports successive improvements leading to the elimination of temporarily malfunctions inherent in the early stages of any innovator process. But the path is open and, surely, the comming versions will bring the optimizations that all of us want them., including us from the NSI. Because, never forget, some of the most important and steadfast users of official statistics, are even the producers of these statistics.

4.   References

PHC application v.0.10 User Guide.