Paper 044-2010

A SAS® Business Intelligence – A Technology for Register-based Censuses

Dr. Rami El-Berry, Christine Hallwirth, MHE & Partner GmbH, Vienna, Austria

ABSTRACT

Any member of the United Nations has to carry out an entire national census between the years 2005 and 2014. While the demographic characteristics of the population in the still very widespread traditional census are determined by directly contacting citizens, the new kind of register-based census represents a radical shift in the way censuses are conducted, involving an entirely new set of tasks and problems.

The basic principle of the new census process consists of the automated and anonymous matching of national registers to implement the comprehensive reporting necessary for each census.

The current article describes an approach to the realization of the new census with the aid of SAS® BI technology, which is based on the OLAP principle and thus covers various topics such as data integration, data quality, statistical matching, and, of course, classical data-driven reporting in this very innovative and dynamic environment.

INTRODUCTION

Between the years 2005 and 2010 an entire national census has to be carried out by every member state of the United Nations. While the demographic characteristics of the population in the (still very widespread) traditional census are determined by directly contacting citizens, the new kind of register-based census represents a radical shift in the way censuses are conducted, involving an entirely new set of tasks and problems.

The basic principle of the new census process consists of the automated and anonymous matching of national registers to implement the comprehensive reporting necessary for each census.

The current article describes the Austrian approach to the realization of the new census with the aid of SAS® BI technology, which is based on the OLAP principle and thus covers various topics, such as data integration, data quality, statistical matching and, of course, classical data-driven reporting in this very innovative and dynamic environment.

2010 WORLD CENSUS ROUND

Within the framework of the 36th meeting of the UN Statistical Commission, a worldwide population and housing census was decided. This census, which provides one of the most important bases for decision making for all further planning in the overall interest of the state, is to be carried out by every member state at least once between the years 2005 and 2014 [1].

A new development, "register-based" census, is increasingly taking over from the traditional, generally paper based, population census with the purpose of easing the burden on the respondents, reducing costs and increasing data accuracy and consistency. In the past, citizens answered questions directly, however, in the register-based census, a number of administrative data sources shall be taken into account in order to be brought together in accordance with the required data protection regulations. The necessary legal foundations in most European countries, not least in Austria and Germany, are based on the UN Census Recommendations [2], respectively on the version adapted by the European Union [3]. For all countries belonging to the European Union, this regulation is considered obligatory whereby a standard comparable basis for the data should be created.[1]

The general nature of this new register-based census shall be explored using Austria as an example.

[1] Since the EU Census Recommendations is very close to the UN Census Recommendations, those referred to with the term World Census Recommendations (cf. [2], page 3), such census results, for example population, but also families, households, employment and much more, can consequently be compared at a worldwide scale.

AUSTRIA AS AN EXAMPLE OF THE NEW CENSUS

Like in many European countries, a real register-based census will also be carried out for the first time in Austria in 2010. A law has been passed in order to check feasibility and to develop the necessary concepts [4]. In particular, it is about evaluating whether data demands regarding the necessary features in the sense of the "Core Topics" of the Census Recommendations can be derived from the administrative register or whether further legal or organizational principles have first to be created. Therefore, a complete register-based census was arranged and carried out in full with a deadline of October 31, 2006.

Since the same features, such as date of birth for example, are naturally included in a number of different registers, the principle of redundancy was used to identify any inconsistencies. This resulted, not least, in the residency analysis which detects obsolete data [5], page 8.

As a basic principle, the Austrian administrative community assumes that a real living person has to appear in several registers. For example, a person over the age of 18 and living in the national territory must appear in at least one other register in addition to the central register of residents, such as the social security register or the register of the public employment office, for example. The same applies for persons under the age of 18 who, for example, when they turn 6 have to be listed in the register of enrolled pupils as well as in the register of residents. In Austria, a total of over 40 different registers were joined.

Creating a standard key was also a particular challenge. The traditional individual registers which are already relatively old in some ways, could not just be easily interlinked with one another because there was no common key. This demand led to an anonymous person key being introduced by law. Within the framework of the trial register census, the individual register-led bodies developed methods in order to generate this anonymous key, which is based on the register of residents, for their respective databases. Then they transmitted the data resources enriched with the respective compulsory legal characteristics to the Central Office for Statistics ("Statistik Austria"). Each Central Office for Statistics was, and ultimately is, the institution which carries out the actual core work of the register-based census and interprets and publishes the results.

THE CENSUS IN EUROPE IN GENERAL

Apart from France, who already carried out the census in 2006, a census will be held in all other European countries in 2010 or 2011, whereby many countries are attempting to carry out a register-based census (partly with and partly without the traditional accompanying surveys). The northern European countries (Sweden, Norway and Finland) seem to have made the most progress which reflects many years experience with register surveys. On the other hand, in Central Europe, including Germany[2], a register-based census is a new thing.

TYPICAL DEMANDS WITHIN THE FRAMEWORK OF A REGISTER-BASED CENSUS

Within the framework of a register-based census project, there is a number of technology and content based issues which typically arise and are generally also independent of the national, specific situation.

The documentation and traceability demands are portrayed as the top priority as they are typically requested in the underlying law. In concrete terms, this means that the resulting individual accompanying characteristics from the census result must also always be traceable back to the sources.

A further important point can be seen in the problem of merging records from various registers which are meant to be the same, i.e. from the same person, and have no anonymous key. This procedure is often referred to as "Record Linkage" or "Record Matching" and can be seen in countries which have introduced central anonymous person keys (cf. e.g. Austria [5], page 36).

Normally, the central analysis work within the framework of such a project is carried out by specialist users for the respective areas such as employment or housing, after which the overall result is compiled and checked once again for consistency and plausibility.

A result of the overall process is often a better insight which leads to a single block level in a correction procedure.

THE SAS® BI PLATFORM IN THE CENSUS ENVIRONMENT

The default installation of the SAS® BI platform already has lots to offer in terms of functionality, which is actually needed in order to adequately reflect the requirements mentioned briefly above.

[2] For example, Germany will carry out its first register-based census in 2011, which is accompanied by a sample acting as a corrective.

The higher demand for documentation and traceability in the procedures which implement the data flow can easily be met through the use of the SAS® Data Integration Studio. Generally, when the SAS® Data Integration Studio is used on a large scale basis, the continuous usage of the tool is a prerequisite for the successful integration into complex projects. Therefore also "User Written Code" (elements which contain program code formulated in Base SAS®) should be avoided wherever possible. Although a lot of functionality for the proper integration and analysis of manual processing steps has been introduced in SAS® 9.2, experience has shown that the usage of functions like “cause and effect analysis” of the SAS® Data Integration Studio simplifies the overall data flow management and administration.

There is a small but very important specific characteristic in the field of data management within a register-based census - namely the problem of "Record Linkage". Although this topic would need more explanation, not more than some brief comments concerning the technical realization within the framework of an SAS® 9 platform solution shall be given here.

By "Record Linkage" we understand the process of bringing together records which are linked in terms of content but are not technically linked. This can have many causes and can be due to the correct spelling of names or addresses of a record in different registers. Implementing "Record Linkage" requires a feasible concept with regards to the actual content which is linked, as well a study to ensure data validity and a maintainable technical solution for data handling.

As long as the relevant requirements are fulfilled, a solution can be achieved with the help of the DataFlux® dfPower Software®. However, it is often the case that in most organizations which have been using SAS® for many years, corresponding comparison programs are already implemented for specific and regionally very different tasks. The integration of these solutions is sometimes a smaller burden than a complete redevelopment within the framework of an equivalent overall process developed with the help of SAS® Data Integration Studio. However, this decision is to be taken depending on the individual case and cannot be generalized.

As such derived records in the official overall reports created by the relevant bodies are considered as showing a corresponding uncertainty in any cases and therefore can be marked accordingly, the traceability of individual data which is not completely possible can be considered with more leniency within the framework of the already mentioned “cause and effect analysis”.

According to the authors’ experience, it is often the case that, although the determination of the various data operations which are needed for proper data consolidation can be very complex and time-consuming, the final rules that are implemented are often relatively simple and can be easily implemented with the help of the SAS® Data Integration Studio.

The results of register-based census projects are always to be considered in a certain historical context and must be analyzed on the validity of breaks in time series for example; therefore it is important to prepare the accompanying report about the overall term without delay.

Therefore, there are a number of reasons why OLAP – technology, together with the add-in for Microsoft Office, plays a key role in the project’s success. First of all, the existence of OLAP cubes gives expert users an advanced tool to work with. With the help of this tool, the user is in the position to carry out independent plausibility analyses which otherwise would have to be completely defined and commissioned.

Once the associated cubes are defined for every subject area (e.g. employment statistics, housing census etc.), they too can be automatically created with every update of the database whereby the user can observe and control the progress. The SAS® Add-In for Microsoft Office can therefore be integrated into an Excel workbook and play out its strengths regarding data visualization and reference recognition. After integrating a cube enquiry into a workbook in order to compare specific characteristics in the current census analysis in a historical context, the corresponding output of the enquiry is updated automatically as soon as the actual cube is recreated. The conditional formatting very quickly portrays reporting solutions for the quality control in combination with user defined SAS® formats.

An area within a register-based census which is typically very important is the above mentioned data correction process, where errors or omissions identified in the data can be correctly created with “manual” or semi-automatic keys which can be applied to the entire data body.

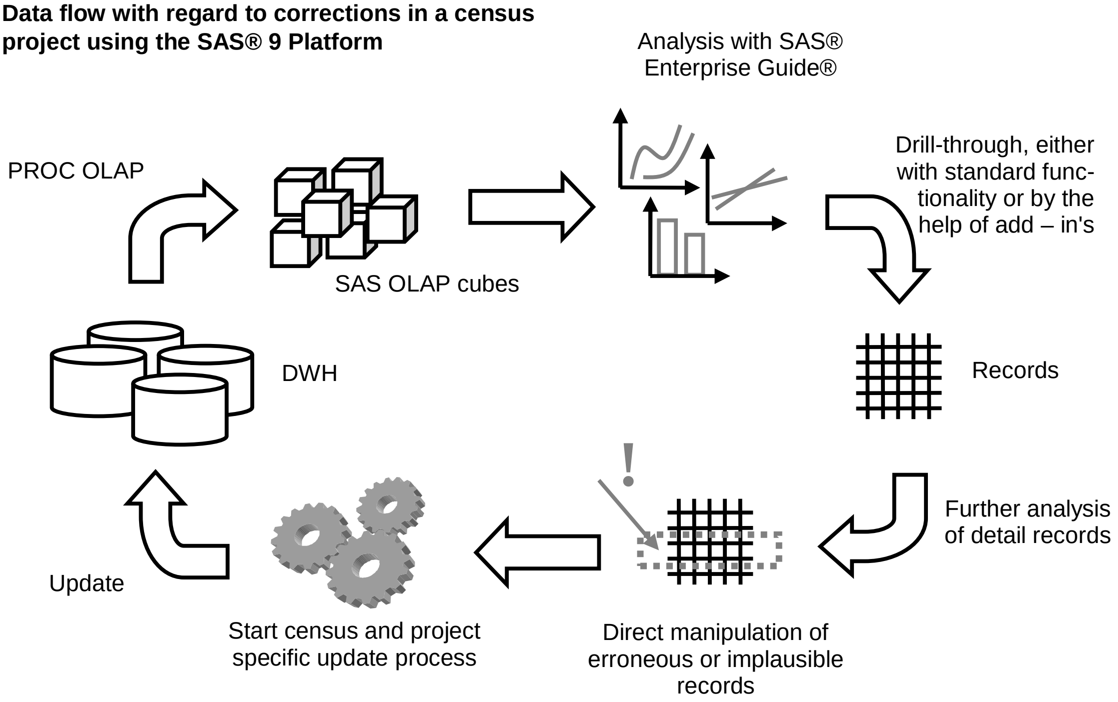
Although a direct update of the OLAP cubes is not currently planned in the platform[3], and a direct manipulation of the data would also go against the nature of this administrative process, the OLAP technology supports the overall process well by means of:

[3] Within the SAS® Financial Management, these functions were implemented for a specific field of application.

- passing through detailed data and/or
- add-in expansion possibilities of the SAS® Enterprise Guide®.

When passing through detailed data, it is possible to request the concerned records in the form of a SAS® Dataset which portrays the currently displayed enquiry in its entirety. This mechanism is subject to certain restrictions with regard to data quantity which can be usefully retrieved. The records retrieved in this way can be analyzed further and, if necessary, restored. This can be done by a corrective mechanism with a typical tool such as the DI Studio, in order to ultimately be available for the next planned recreation of the cube.

Data flow with regard to corrections in a census project using the SAS® 9 Platform


Analysis with SAS® Enterprise Guide®
PROC OLAP
SAS OLAP cubes
Drill-through, either with standard func-tionality or by the help of add – in's
DWH
Records
!
Further analysis of detail records
Update
Start census and project specific update process
Direct manipulation of erroneous or implausible records

The add-in expansion mechanism of the SAS® Enterprise Guide® enables it to carry out a detailed enquiry of data that are enriched with additional variables related to content. What sounds relatively theoretical has a significant practical use.

An expert user typically analyses specific distributions of family and household classifications for plausibility. E.g. if a certain percentage of households containing several people seems unreasonable and is therefore technically not justifiable we have a closer look. If an imbalance arises here, it can be of great use, in addition to the existing view of the cube, to retrieve all such records which are located near to those considered and, however, no family or household is assigned. This process can be available for the user via a simple click in the relevant add-in menu of the SAS® Enterprise Guide®. The developed add-in must request the MDX Statement compiled in the background by the SAS® Enterprise Guide® which represents the user's current view of the cube and can then request the corresponding individual records via a "PROC SQL" Request to the SAS® Workspace Server. The table retrieved, filled with the desired information, like all records not yet assigned to the same location from the example, can be directly integrated into the user's SAS® Enterprise Guide® flow chart by the user him/herself as a SAS® Dataset, like what he/she is used to with the standard components such as "Filter and Query...", for example.

CONCLUSION

Over the next few years, most of the states will update their procedures in the course of the upcoming census round. A considerable number of them will thereby turn their backs on the traditional census and go for a purely register-based one.

This poses a series of challenges for the underlying IT infrastructure as well as for the solutions in place for data management and data analysis. We have described those points which we believe to be the most important based on our project experience of European censuses and we have briefly touched upon the corresponding problems and showed how a large part of this could already be efficiently covered with the standard functionality of the SAS® BI Platform.

The proposals for an effective solution are in principle based on an OLAP-centered display format in the course of the interaction with the relevant expert users and the SAS® Add-In for Microsoft Office as well as SAS® Enterprise Guide® as intelligent tools for the portrayal and analysis of the mostly very large administrative data body which typically arises in the course of a large-scale census.

REFERENCES

[1] United Nations, Statistics Division: Documents for the thirty-sixth session of the Statistical Commission New York, 1 to 4 March 2005 , http://unstats.un.org/unsd/ statcom/sc2005.htm

[2] United Nations, Statistics Division: Principles and Recommendations for Population and Housing Censuses, Revision 2, New York 2008

[3] United Nations Economic Commission for Europe, Statistical Office of the European Communities: Recommendations for the 2010 Census of Population and Housing, New York and Geneva 2006

[4] Act for the register-based census in Austria ("Österreichisches Registerzählungsgesetz"), BGBl. I Nr. 33/2006, March 16 2006

[5] Central Office for Statistics in Austria: Report on the test register-based census 2006 ("Bericht über die Proberegisterzählung 2006"), Vienna 2009

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Christine Hallwirth, Dr. Rami El-Berry
Enterprise: MHE & Partner GmbH
Address: Fischhof 3
City, State ZIP: A-1010 Wien, Austria
E-mail: office@mhe-partner.at
Web: www.mhe-partner.com