From Data to Information: Exchanging and Analysing an Extensive Dataset Between Different Software Packages

CAQD 2012: 14th Conference on Computer-Aided Qualitative Data Analysis [2012.03.07-10, Marburg DE]

Francisco Freitas [franciscofreitas@ces.uc.pt] | Centre for Social Studies - University of Coimbra [www.ces.uc.pt]

Context

Data collected under a research project funded by the *Fundação para a Ciência e Tecnologia* (FCT) [PTDC/CS-SOC/099134/2008] and coordinated by the Centre for Social Studies (CES). This project is devoted to the analysis of Participatory Budgeting experiences in Portugal and Cape Verde. In total, 9 case studies are currently under way.

A Work Package

A variety of research activities has been conducted so far. One of the tasks was surveying and evaluating the 4th Cycle of Participatory Budget (PB) in Lisbon in 2011. The Municipality of Lisbon has decided to assign this assessment/audit to an external academic institution. To that end, a **web survey** has been published.
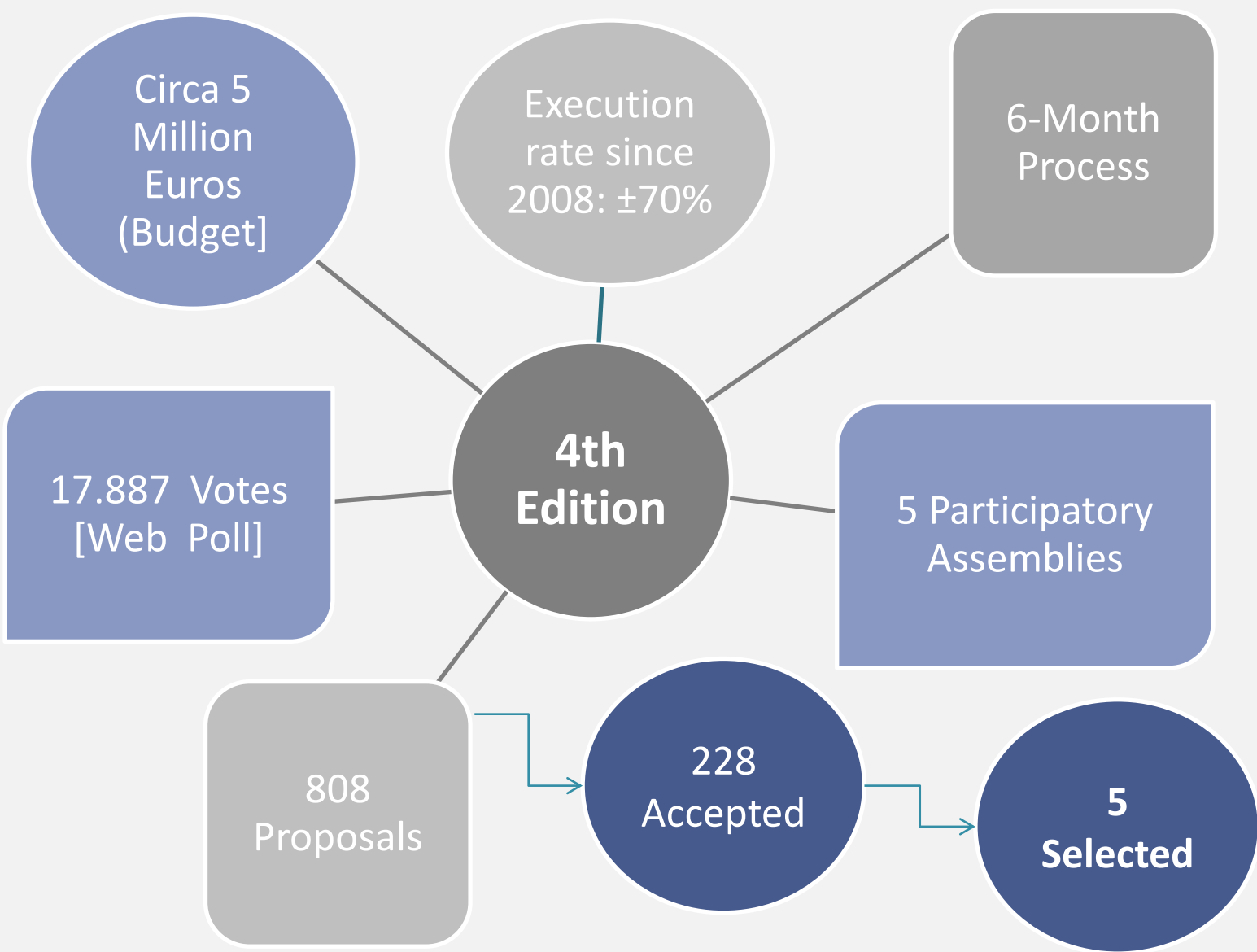
Web Resources

The Municipality of Lisbon developed a special web portal entitled *Lisboa Participa*:
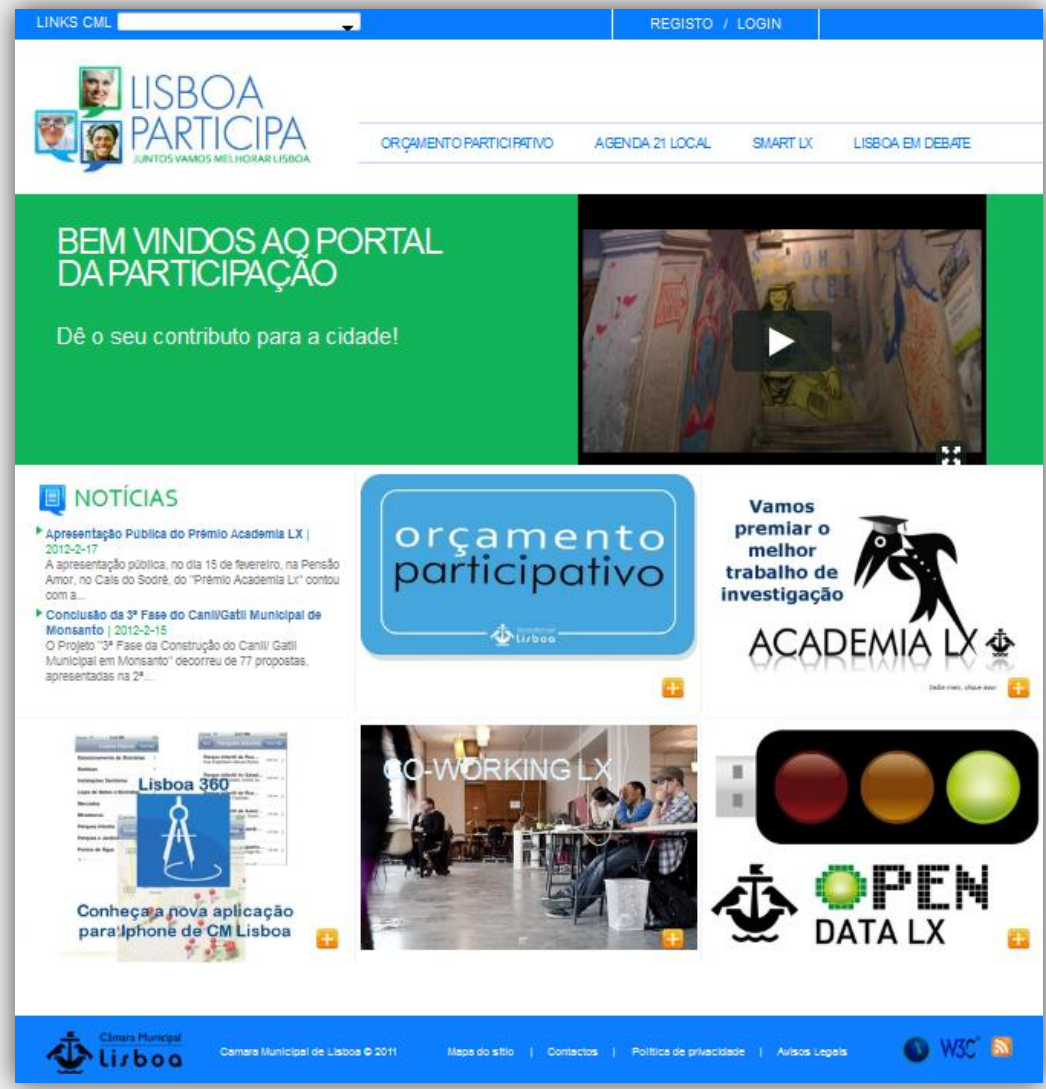
Citizens are kindly invited to visit this web portal and then grant access to very relevant information in an range of domains. They are as well allowed to contribute with their views and opinions in some fields. Among many other activities, through the portal it was possible to submit and vote proposals related to the PB.

Lisbon PB in Numbers

Circa 5 Million Euros (Budget)
Execution rate since 2008: ±70%
6-Month Process
17.887 Votes [Web Poll]
4th Edition
5 Participatory Assemblies
808 Proposals
228 Accepted
5 Selected

Survey Research

The survey was prepared by a CES research team. Several meetings were conducted with local officers in order to improve the former versions of the questionnaire. By the end of the cycle, a link was published in the portal, directing citizens to the **scientific inquiry**.

>Target Group

According to the last population census, there were 564.657 inhabitants in Lisbon in 2011 (INE, 2012). The PB general regulation allows full age citizens who live and/or work in Lisbon to participate in the activities. The process puts a strong emphasis on **electronic participation**.

Nevertheless, a set of assemblies are organized every year in order to facilitate the contribution of citizens who prefer to attend in-person meetings. Also, four offices with dedicated staff are available to help citizens voting.


27.046 Registered Users

Sample (n)

2.855 Respondents
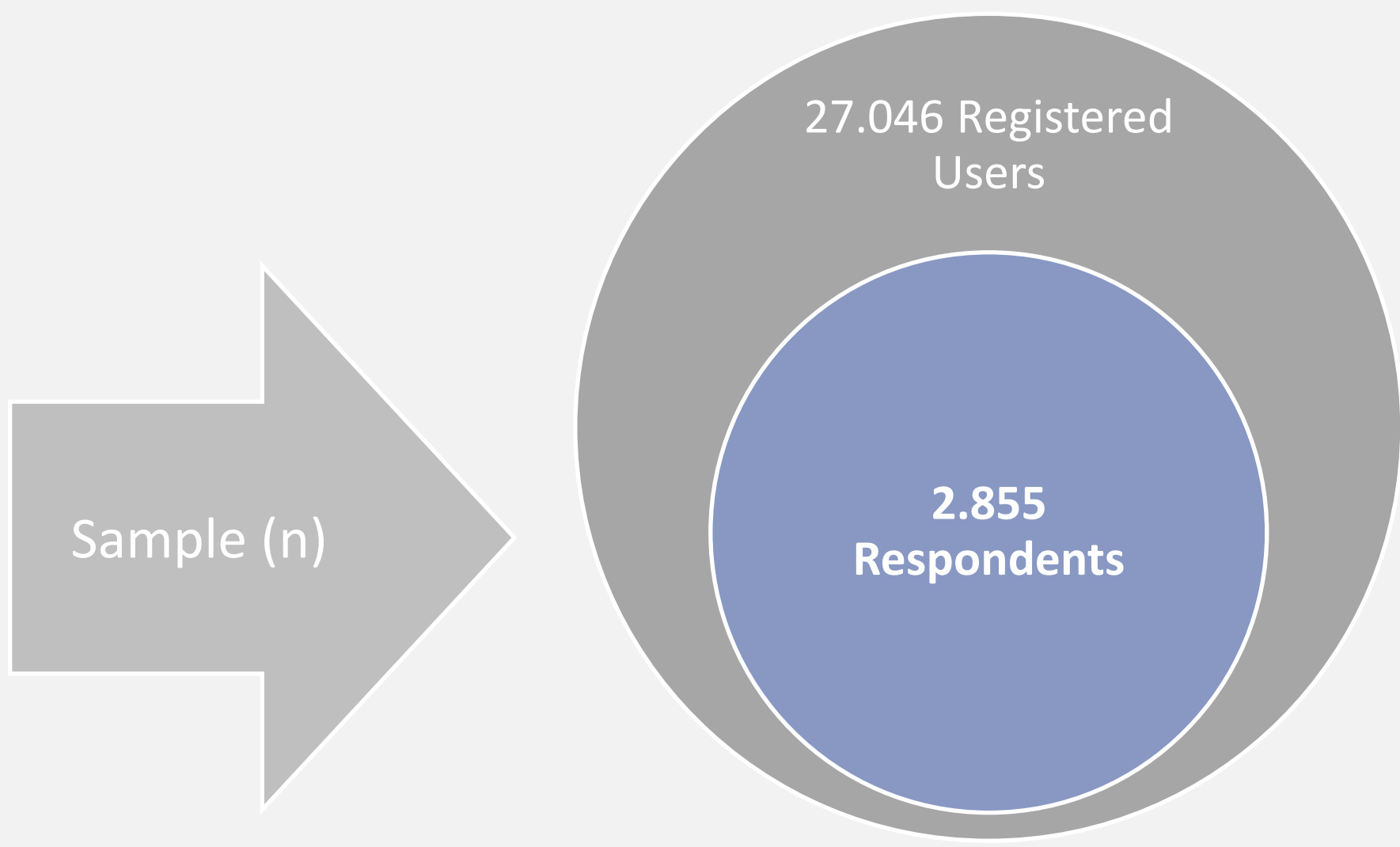
>The Electronic Questionnaire

|	
Total Number of Questions	54
Array Filters [i.e. optional subsequent questions]	13
Short Text Entry Boxes [i.e. other option(s)]	20
Open-Ended Questions	2

>Web Hosting

The appropriate web link was embedded in the *Lisboa Participa* web portal. The questionnaire was published using *LimeSurvey v1.90* (LS). The software was hosted on a specific server at CES. Below there are images of the program's dashboard and a survey's welcome screen:
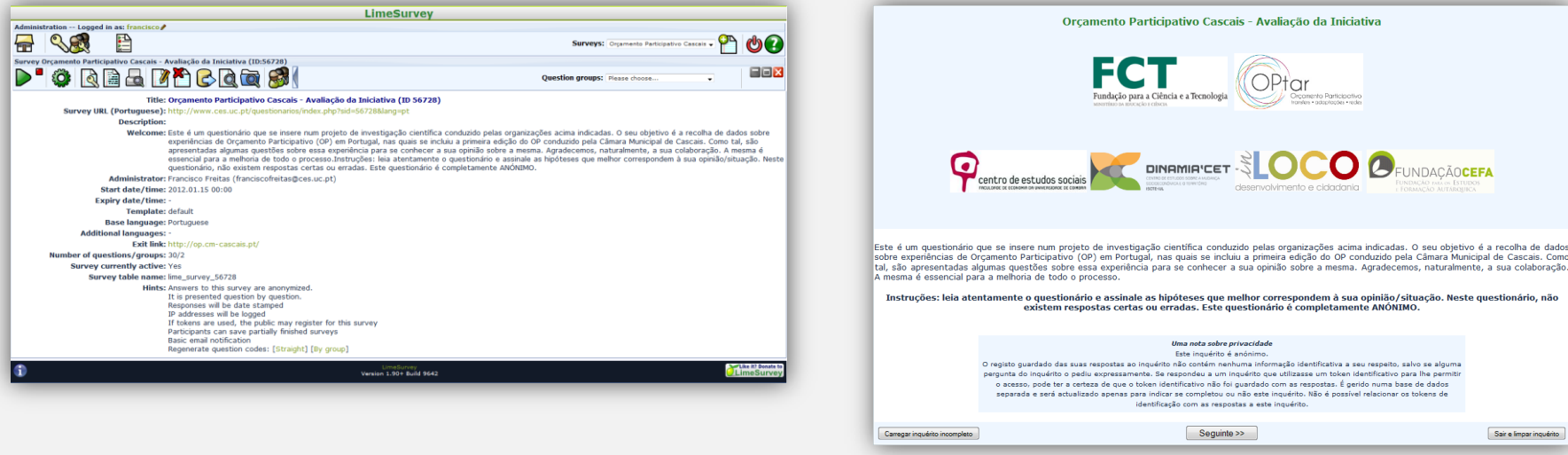
Managing the Data

>Quantitative 'Assets'

Exporting the collected data in LS can be done in different ways. In this case, the whole dataset was exported to *IBM SPSS v19* as many variables required statistical analysis. The user finds a command to perform such action in LS's dashboard. Two files are generated: a *.dat* file, containing the survey data, and a syntax file (*.sps),* where all the variables are defined. The latter is quite useful as the user is able to program and transform the strings according to the needs:
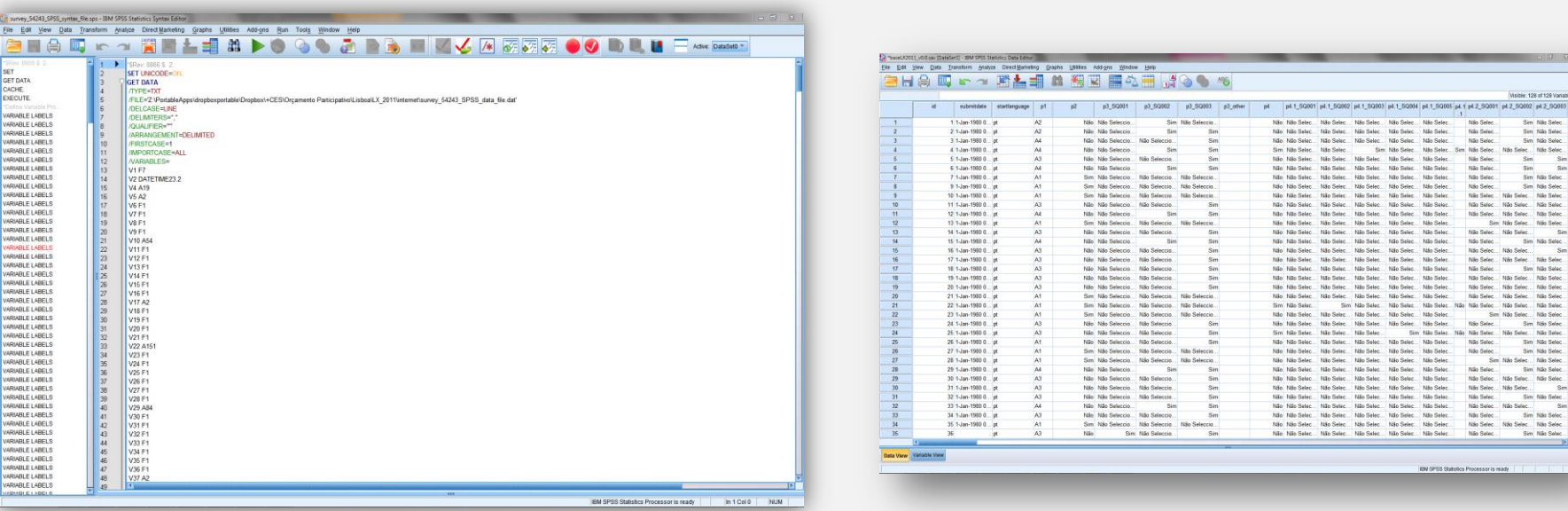
>GIS Integration and Spatial Analysis

Thematic cartography is very useful to represent changes over time and to promote comparisons between different zones. In that respect, some variables have been georeferenced in order to broaden the scope of the analysis. This task has been performed using *Quantum GIS 1.7*. Two examples are displayed afterwards:
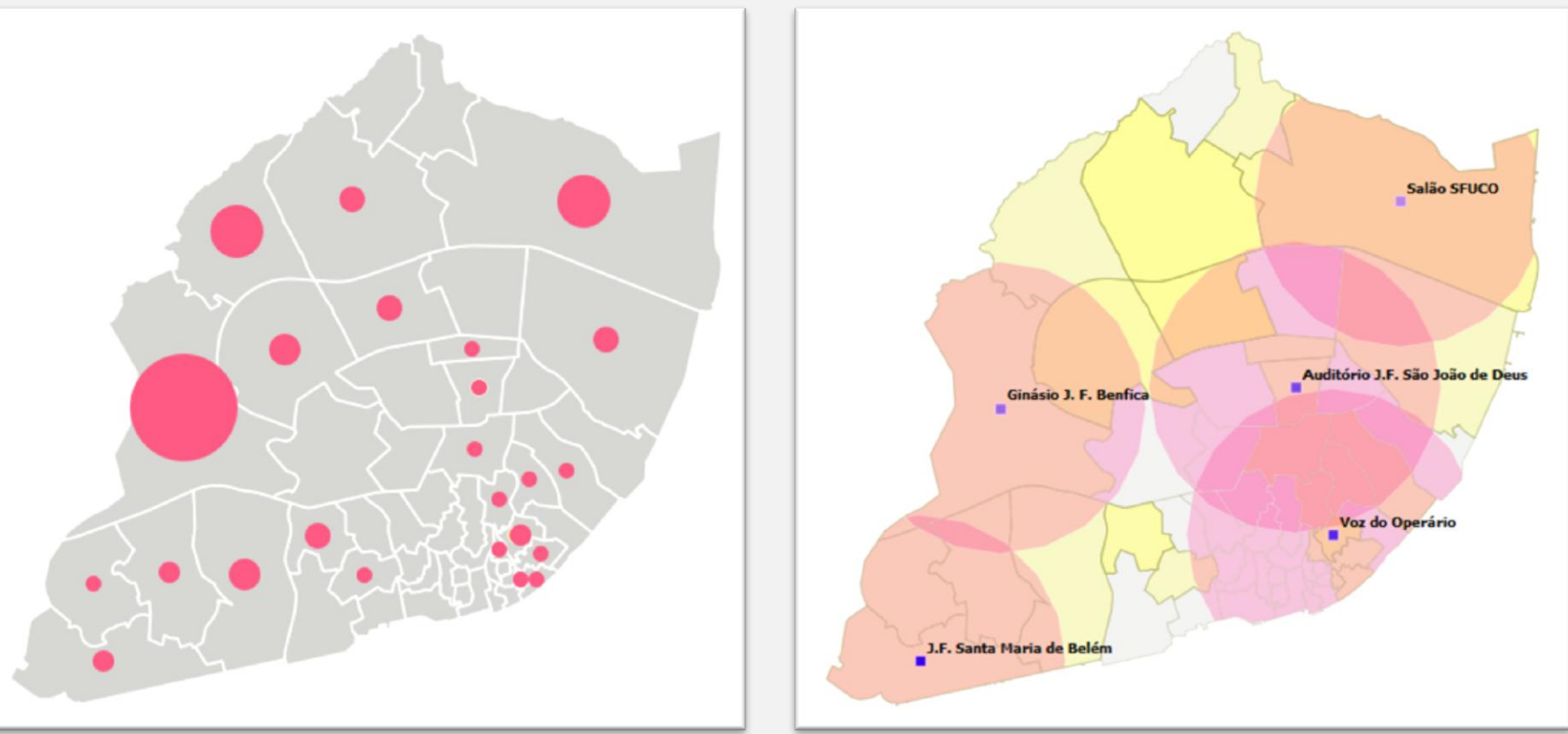
>Qualitative Variables

This type of survey implies numerous texts and not all are necessary short in size – some respondents do like to express themselves. The time-demanding text transcription is therefore avoided, but other challenges emerge.

MAXQDA Relevance

_Preprocessing the Data

Considering the sizeable amount of string type variables, the usage of a CAQDAS package would be an advantage. Extracting the appropriate data from a *SPSS* file is a very straightforward operation, allowing the user to set very relevant parameters.

An *Excel (.xlsx)* spreadsheet was therefore employed to bridge the data into *MAXQDA v10*. Spreadsheets prove to be useful in rearranging data files. In this case, is was only necessary to produce minor adjustments in the data table in order to be intelligible by *MAXQDA*. That means adapting some fields and using new labels like 'textgroup' or 'textname':
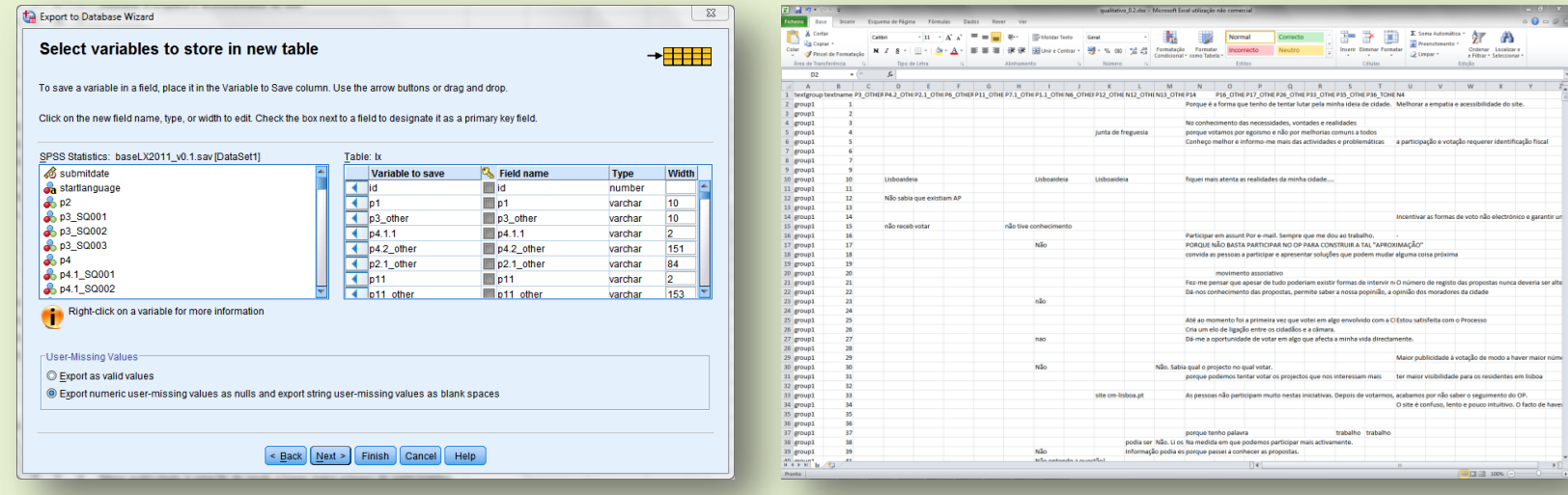
Once the file is ready, it is possible to import it directly into a *MAXQDA* project file *(.mx4)* using a dedicated command. The described steps are part of the **Preprocessor** engine.

During this operation, an important step is to avoid importing empty texts, which correspond to blank fields in survey response. They are of no use because every single text segment is identifiable by a questionnaire number and by a variable name:
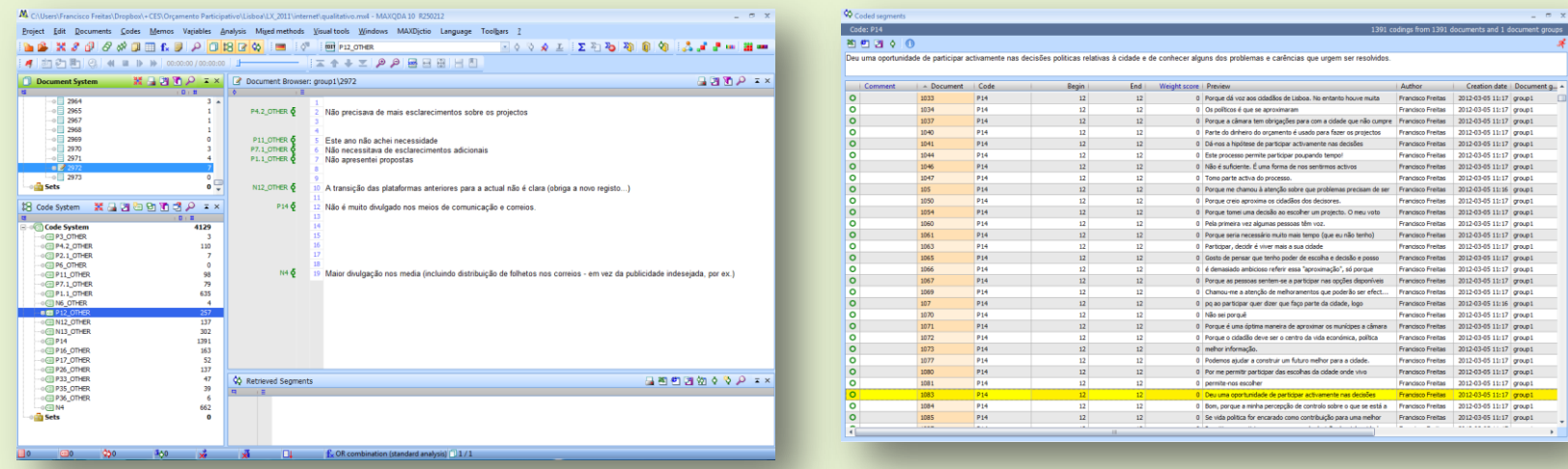
_MAXDictio

Autocoding must be taken into consideration when a researcher is handling a few thousand text entries. In this case, the *MAXDictio* add-on is the tool that has been used to foster the process. After developing a categorization and after fine-tuning the dictionary, it was possible to group the answers in a fruitful way.

Final Remarks

❑ Electronic questionnaires are a powerful tool that can optimize the data collection process in many ways

❑ Correct survey design remains a key element

❑ There is an (empirical) reason to believe that the easier to collect the data in surveys, the more will be asked to respondents…

❑ Increased rates of (voluntary) participation in surveys are challenging as the collection of qualitative data can progress to a level where is difficult to avoid a deceptive or inaccurate analysis of data

❑ These digital endeavors do not discard the role of researchers but simply demand for different skills (i.e. from a 'paperwork officer' to a an effective data handler, manager and analyst)

❑ The communication and exchange between different packages is facilitated if certain procedures are adopted (e.g. univocal ID's)

❑ Extensive datasets are of no use if no relevant information is attained. In that respect, software packages play a major role, saving resources

References

Available under request, please contact the author.