Distr.
GENERAL

Working Paper No.17
14 March 2008

ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Meeting on the Management of Statistical Information Systems (MSIS 2008)
(Luxembourg, 7-9 April 2008)

Topic (iv): User perspective

VISUALISATION OF NEIGHBOURHOOD STATISTICS USING GOOGLE EARTH

Invited Paper

Prepared by Olav ten Bosch and Edwin de Jonge, Statistics Netherlands, the Netherlands

I. INTRODUCTION

1. Many statistics of Statistics Netherlands are available at neighbourhood level. These figures were already accessible via the statistical database, *StatLine*, but they can be presented more attractively with modern geographical web-tools used on other websites. For example, in the Netherlands more and more people look for houses using websites of real estate agents that project houses for sale on *Google maps*. Information on traffic intensity and crime prevalence can be looked up via popular websites that display this detailed information on Google maps. People looking for used bikes, cars, furniture or computers can even see what their neighbours have to offer using a geographical service of a well-known website for exchange of second-hand goods called "marketplace". We think neighbourhood statistics can be disseminated more naturally if they are presented in the context of tools and media people use in everyday life.

2. Therefore, Statistics Netherlands has started to publish detailed neighbourhood statistics in combination with a projection of neighbourhood borders using *Google Earth*. Users can download a statistical layer from the main website of Statistics Netherlands. The streaming mechanism of Google earth is used to guarantee a high performance service, in which a large volume of statistical information is delivered to the client PC as part of other geographical photos and data. Links were added to web-articles containing statistics at detailed regional level, showing relevant locations on Google Earth. Furthermore, the same statistical data were also made accessible via a dedicated website for neighbourhood statistics using Google maps.

3. This paper explains the approach taken by Statistics Netherlands in more detail. It highlights some of the choices made and demonstrates the use of these services by a number of examples of these data. Also, it touches slightly on some of the more experimental features we plan to develop for disseminating regional statistics on Google Earth, such as projections of one or more statistical variables on "heat maps" and the possibility to visualise statistics over time.

II. THE CHALLENGE OF PRESENTING NEIGHBOURHOOD STATISTICS

A. Traditional mapping versus streaming photo-based mapping

4. Statistical offices usually have a huge amount of statistical information at a detailed regional level. And this type of information will increase in the future as various administrative sources are combined. In the Netherlands, we call these statistics *neighbourhood statistics*. Traditionally, neighbourhood statistics are presented in the form of tables and maps. This works well for people who regularly visit Statistics Netherlands' website and know their way to the mapping module. They typically have the skills and patience to browse the immense datasets in the statistical output database *StatLine* and generate the maps of their choice. Using this mapping engine maps can be generated as-you-go. Figure 1 shows two examples of maps generated from the statistical database *StatLine*. The example shows maps at municipality level, but Statistics Netherlands has more detailed levels.

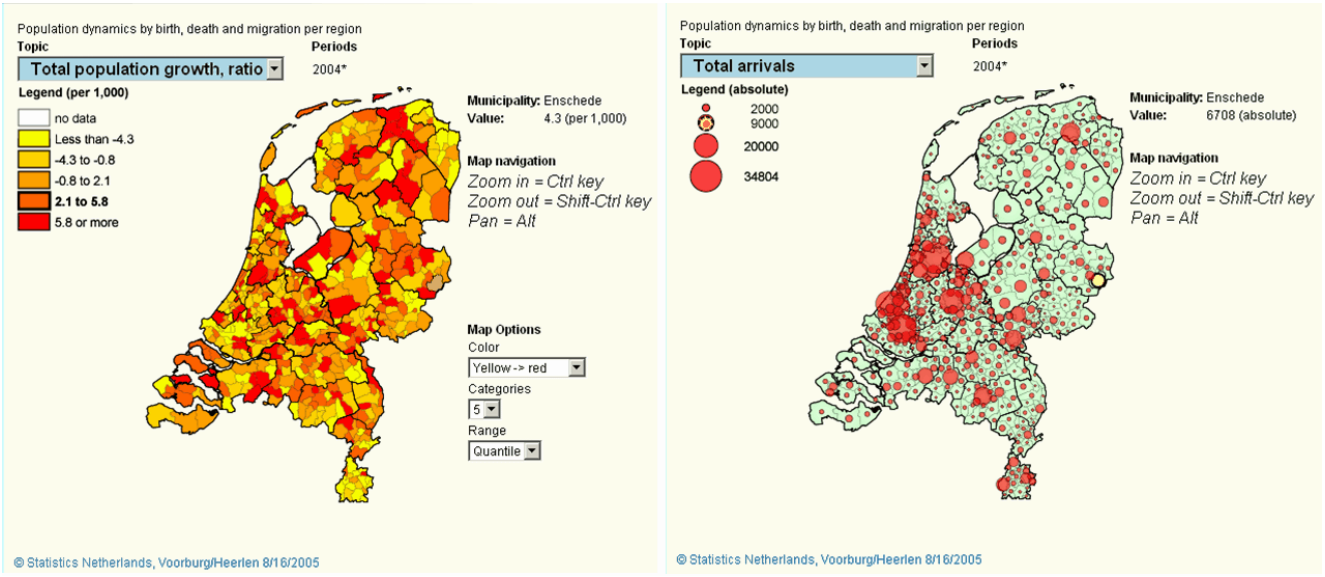

Figure 1. Municipality maps generated by StatLine

5. Although this type of mapping engine is a useful tool for frequent visitors, generating such a detailed map from the huge statistical database can be a difficult task. First of all, not all data in the statistical database are suitable for presentation in a map. Only data that use one of the standardised geographic dimension formats can be visualised in a map. Secondly, the mapping engine generates maps in scalable vector graphics (SVG). This technique is technically advanced, but most people don't have the appropriate plug-in installed on their computer, which creates another threshold. A third disadvantage of these mapping engines is that users must tune the maps generated via a number of parameters. These parameters may complicate the mechanism for ordinary users. So, in practice there are a number of problems with this type of *traditional mapping engines*.

6. Today many people use map-based applications for common tasks, such as checking for traffic jams or local weather forecasts. If they want to buy a used car, a house or book a meal in the nearest good restaurant, people in the Netherlands can choose between a number of popular map-driven websites. In general, then, people *are* familiar with mapping engines used by popular Dutch websites for everyday tasks. Usually, these websites project their information on maps such as *Google maps* or use more advanced features of *Google Earth*. Figure 2 gives some examples of such sites.

7. These sites use advanced technologies to display these maps smoothly on the users' screens. The maps are created from *satellite photos* or aerial photos that display the area in more or less detail. These photos are tiled in the correct order so that the user has a "flying" experience while inspecting data on the

map. Because of the huge amount of photos needed to display the almost endless possible maps, these services use *streaming technologies*, where only the appropriate photos are sent to the user and these photos are put together on the client's computer. We call this type of map generation *streaming photo-based mapping technology*.

8. Since 2000 there has been a tremendous increase in the use of streaming photo-based mapping technologies in the Netherlands. One cause of their success in the Netherlands may be that the displayed images are very detailed, as aerial photos were used more and more. Of course, this drastically improves the quality of the services, as people recognise details from their own neighbourhood and their own house or even their own car. This is exactly the reason that these technologies are such a great instrument for disseminating neighbourhood statistics.

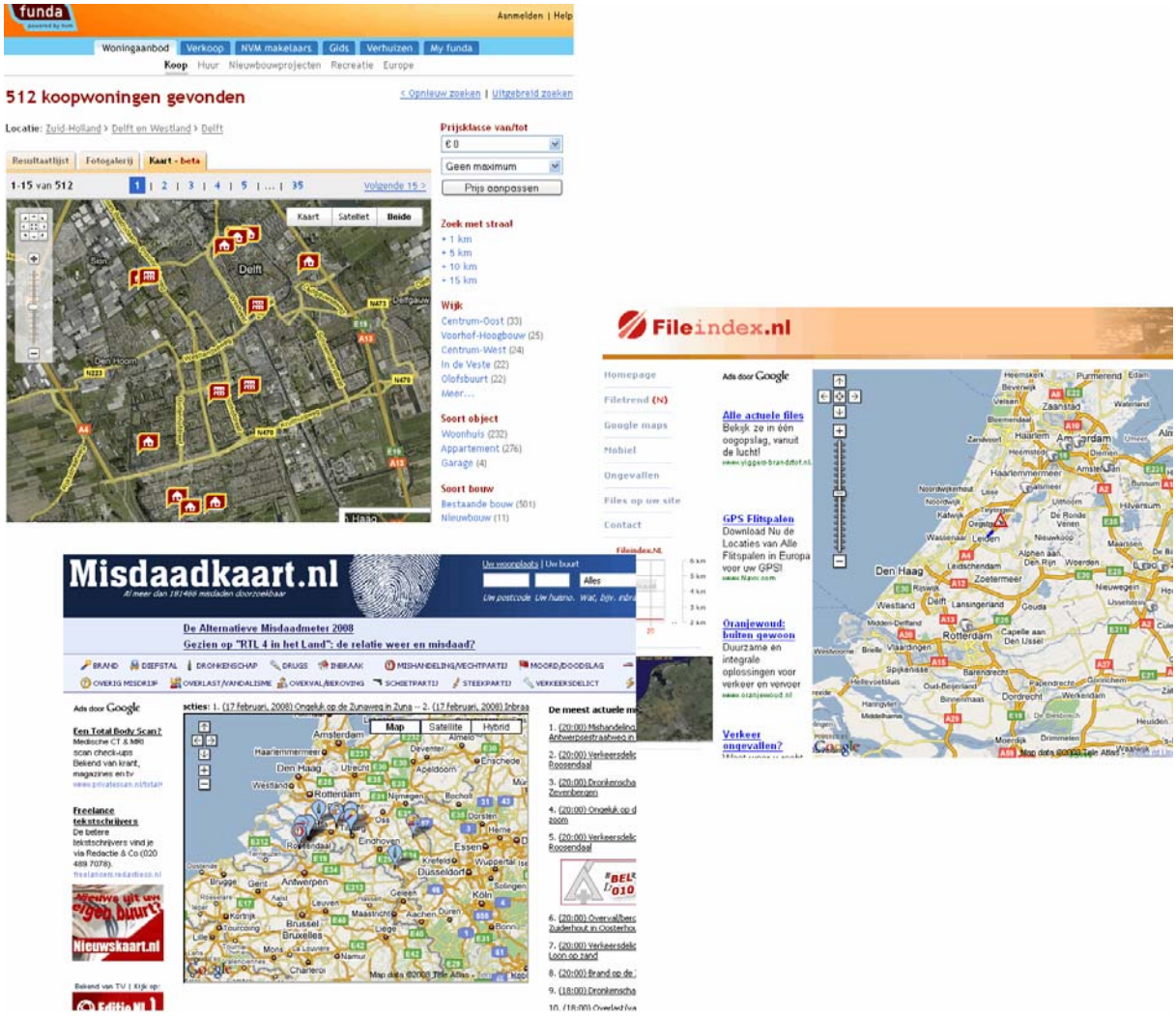

Figure 2. Examples of Dutch websites using Google Maps, houses for sale, crime map and traffic jams

B. A new approach

9. Combining the fact that Statistics Netherlands has very many detailed data up to neighbourhood level that are potentially of very much of interest to the common public with the increase in the use of streaming photo-based mapping engines to project everyday information to the same common public, the use of these technologies for statistical dissemination is an obvious next step. From the viewpoint of usability, what could be better than presenting people with statistics in the context of mechanisms with which they are already familiar.

10. The approach Statistics Netherlands took in 2007 was to use modern streaming photo-based mapping tools for the presentation of neighbourhood statistics. First of all, we tried to find out what kind of information should be included. People were interviewed about what they would like to know about their own neighbourhood. Based on their answers a subset of variables was taken from the statistical database *StatLine*. This resulted in a dataset of about 20 variables for about 11,000 neighbourhoods: ranging from population characteristics such as age, household composition etc., to income and employment.

11. With respect to the technology to be used there were only a few alternatives. Microsoft had *Microsoft Virtual Earth;* Google has its *Google map* technology and the more advanced *Google Earth;* and an open source initiative called *Open Streetmap* was also available. Experiments showed that the Google services were technically most suitable for our purposes. But more importantly, looking at our intended audience (the Dutch public) we saw that the two Google technologies were far more popular than others.

12. The resulting service is a combination of a website, called “CBS in your neighbourhood” (see http://ww.cbsinuwbuurt.nl (in Dutch only)) and a statistical layer on Google Earth. The website uses Google maps to project the statistical variables on a Google map. This website came in production on 12 February 2008 and attracted 20,000 visitors on the first day. A screen dump of the website is shown in figure 3. In addition to the website “CBS in your neighbourhood”, Statistics Netherlands developed a statistical layer on Google earth. This dual approach was necessary to combine ease of use for the public with powerful facilities such as combining statistical information with other layers. The website “CBS in your neighbourhood” serves the first use. The Google Earth layer facilitates more advanced use. The remainder of this paper focuses on the latter approach.

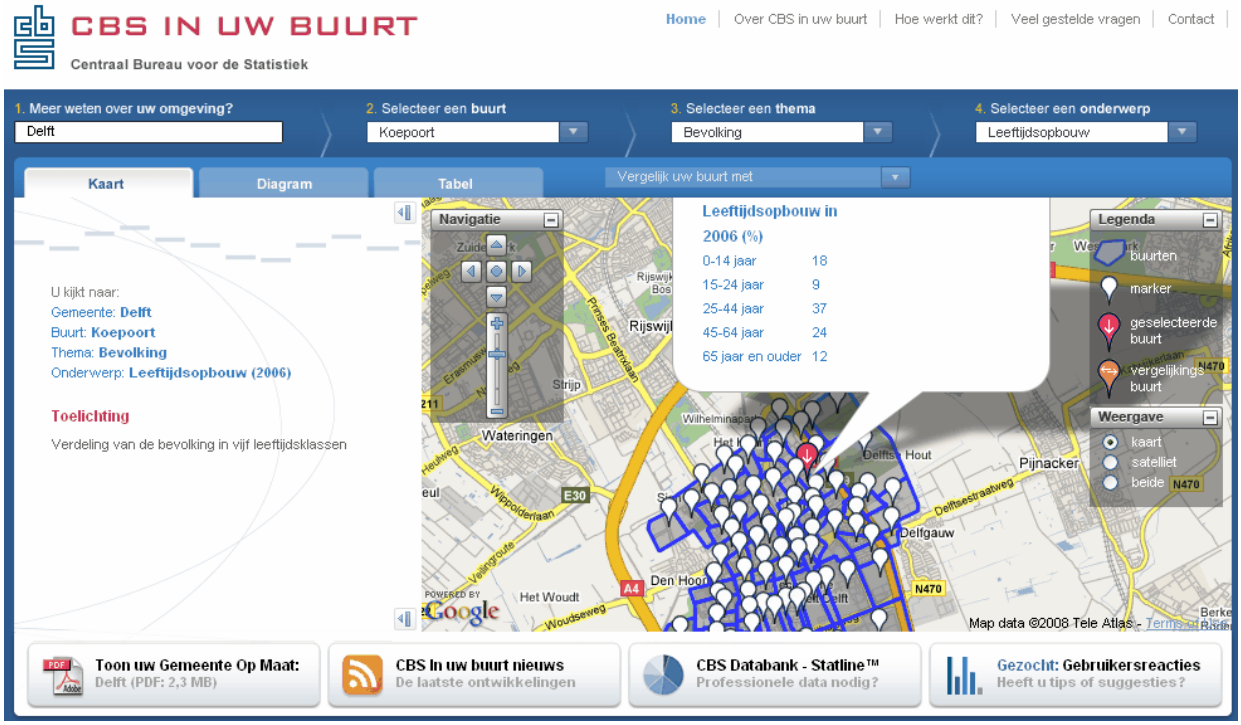

Figure 3. Website "CBS in your neighbourhood"

13. The Statistical layer on Google Earth is offered to the public from the main website of Statistics Netherlands [1]. Users must install Google Earth on their computer to use this layer. Google Earth can be downloaded from Google without charge. Once it is installed, selecting the link on the webpage of Statistics Netherlands will do the job. Google Earth will start up with detailed borders of all 11,000 neighbourhoods in the Netherlands, names of the neighbourhoods and a Statistics Netherlands (CBS) icon in each neighbourhood. This information becomes visible if the user zooms in to a specific region. An example is displayed in figure 4.

Figure 4. Neighbourhood borders and names on Google Earth

14. The Statistics Netherlands icons are clickable, and when clicked will open to show all variables an info window. Figure 5 gives an example for the neighbourhood "Zierikzee binnen de vesten" of municipality "Zierikzee".

Figure 5. Statistical neighbourhood information on Google earth

II. HOW IT WORKS

15. The layer Statistics Netherlands provided on Google Earth uses the underlying streaming technology of Google Earth. This means that data *and* border coordinates are transferred only if the user zooms in on this area. The satellite or aerial photos for this area are transferred from the streaming mapping provider and the statistical data for this area are served from the hosting infrastructure of Statistics Netherlands. The Google Earth client combines these data sources (and potentially many other data sources if the user wants to combine different layers) smoothly into one visual picture.

16. The streaming mechanism is parameterised so that neighbourhood borders and statistical values are transferred "just in time". The size of the statistical data layer including borders is around 60 MB, so streaming it too early would result in performance degradation. The 11,000 neighbourhoods are therefore split into 450 chunks, namely the municipalities. A chunk (i.e. the neighbourhoods of a certain municipality) is only loaded when the user zooms in close enough to a municipality. Experiments have shown that a good setting for streaming these data is to load the statistical data if the municipality occupies more than 450 pixels on screen. Loading the data does not mean that the neighbourhoods will be shown immediately: in many cases this would mean that the screen would be cluttered. The neighbourhood data is therefore shown when a neighbourhood takes more than 150 pixels on screen. This mechanism gives users the feeling of immediate interaction. These settings have shown to be valid for the current dataset, but in coming releases we shall experiment further with different values for different datasets.

17. The data and borders to be displayed in the map are described in *KML* (keyhole markup Language), a language designed by the former company Keyhole, which was bought up by Google. The format has been actively further developed by Google. It is XML-based and has many facilities for displaying markers, info windows, lines, rectangles and even 3D models. We only used a subset of its possibilities, but may use other concepts for coming releases. The statistical data and borders are described in KML files in such a way that Statistics Netherlands can monitor the use of the Google Earth layer very precisely. In principle it is possible to detect which neighbours people are most interested in.

18. Statistics Netherlands publishes a web magazine on a regular basis. The short articles in this web magazine usually contain a hyperlink to the data in the statistical database the article was based on. We used the same approach to insert links to neighbourhood information in relevant web magazine articles. A web magazine article was published that contained links to camera positions on Google Earth. It was thus possible to direct readers directly to look at the most *densely populated* neighbourhood in the Netherlands, or the neighbourhood with the *highest percentage of young people* or the neighbourhood with the *most expensive houses* in the Netherlands. Figure 6 shows a screen dump of this web magazine.

Figure 6. Web magazine with Google Earth links

III. FUTURE WORK

19. The Google Earth layer currently published on the website of Statistics Netherlands is only a first step on the road of new ways of disseminating statistics. For example, we have noticed that the current layer makes it really easy to get a quick view of the characteristics of a specific neighbourhood, but it does not help users much to get a whole view of the status of a certain variable for a municipality, district or the Netherlands as a whole. This would require a more advanced mechanism possibly using *heat maps* to display the status of the Netherlands in one view. The figure below gives an idea of what this may look like. It displays population density from low (blue) to more densely populated (yellow) to high (red). Early experiments with this facility are promising.

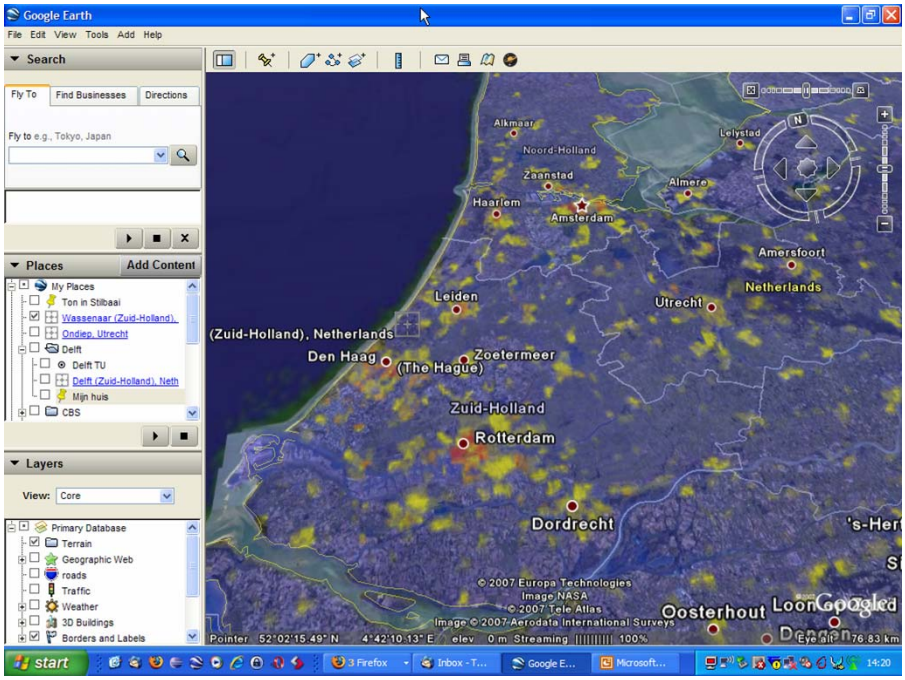

Figure 7. Population density in a heat map

20. Another promising area is the addition of *animated* statistics to the Google earth layer. Recently, Google Earth added a facility to add time to layers, thus making it possible to visualise statistical trends in the context of streaming photo services. This could be used to show changing neighbourhood borders or changes in municipal population over time. We shall experiment with this in the near future.

IV. CONCLUSION

21. Statistics Netherlands has taken a new approach to disseminating neighbourhood statistics. Instead of using traditional mapping techniques, it has introduced streaming photo-based mapping techniques such as Google maps and Google Earth. A website, called “CBSinyourNeighbourhood” that uses Google maps was launched for the public. A statistical layer on Google Earth was published via the main website of Statistics Netherlands for more advanced users. In this way neighbourhood statistics are presented using tools and media people use in everyday life.

22. This statistical layer on Google Earth was implemented in such a way that performance is optimal for users. Experiments with settings for just in time loading have resulted in an optimal configuration where borders and data are loaded just before they could be requested by the user. This “just in time” loading mechanism uses the underlying streaming mechanism of Google Earth. In future we plan to add new elements to the Earth layer, such as projections of one or more statistical variables on "heat maps" and the possibility to visualise statistics over time.