Cartography, Google and neighbourhood statistics*

Note by Statistics Netherlands

Summary

The note focuses on using modern spatial techniques in disseminating statistics. It considers the role of maps in Statistics Netherlands, the opportunities and challenges related to the use of registers for linking data and geography and the recent developments in statistics and cartography. The note also outlines some future perspectives in disseminating spatial statistics.

* Based largely on an earlier paper Cartography with a capital G: Google and more, presented at the IAOS Conference on Reshaping Official Statistics, Shanghai, China, October 14-16, 2008.
I. Introduction

1. Roughly thirty years ago the Dutch novelist Maarten Biesheuvel published a story about two Norwegian friends, Carl and Nyls, who dreamt of creating the ultimate globe. The globe had to be an exact reproduction of the entire world as it was on August 21st 1975 at 12 o’clock. In our statistical world we would call this a reference point. Everything would be visible through a microscope on this extremely accurate globe: people, houses, trains, ships, offices, factories, hospitals, farms, schools, trees, even the cows in the fields. Additionally there would be special annexes in which genealogical registers would be added for all individuals, animals and so on. Their intention was to map and register the world, or in our statistical perception the aim was to compose a statistical Utopia.

2. Although it was an insane job they dedicated their lives to fulfilling this dream. A character named Nwaak finds out about their plan and sends the two friends a package. The moment the package is posted the world turns pitch-dark. Catastrophes take place for weeks, until Carl and Nyils open the package. The package contains the globe they wished for, but it turns out to be more disastrous than they ever imagined. When it is packed, the earth turns dark, and there are devastating earthquakes when postmen touch the globe by mildly crushing the wrapping paper. This globe turns out to be the mother of the earth, it is her prototype. With a microscope Carl and Nyils are able to use the globe as a window to the world. They can now observe everything anywhere and at any time on the entire surface of the earth.

3. Thirty years after this imaginary tale we are well on our way of becoming a contemporary Carl and Nyils. Thanks to new and innovative dissemination techniques it is possible to view the world from virtually anywhere. With only a few mouse clicks Google Earth and Google Maps, amongst others, offer a window on the world just as Carl and Nyils would have liked it. For some topics, such as the weather or traffic, the available information is even live. Statistics Netherlands participates in these new developments with two unique initiatives: Neighbourhood Statistics in Google Earth and Statistics Netherlands in Your Neighbourhood.

4. From a cartographic point of view however there is an interesting dilemma. Thematic maps are a simplification of reality. They are designed to show the geographical distribution of a statistical variable leaving out all irrelevant details. Using and presenting aerial photos as in Google Earth means reintroducing a detailed reality into the map. From Carl’s and Nyls’ perspective this dilemma is not even a ripple on their globe. But from a dissemination point of view it’s something serious to take into account. How Statistics Netherlands dealt with the dilemma and how we managed to use modern techniques in our dissemination processes is the focus of this paper.

II. Statistics and Cartography: early experiences

5. Traditionally, maps are of paramount importance to everyone who is about to set out on a journey over land or water. Orientation is probably the most significant function of maps. Nowadays maps can also play a role in transferring large quantities of data with a spatial component. Maps no longer have to be printed on paper. They can also be virtual or interactive and can be displayed on a computer screen. The aim of maps is to transfer spatial information. Maps can be used to illustrate differences or similarities across geographical areas. Local or regional patterns, which may be hidden within tables or graphs, are often made clear by using a well designed map. As many statistics share the
geographic component, maps are most suitable for a spatial representation of statistical results.

**A. Dissemination of regional statistics through web maps by national statistical offices**

6. Increasingly, regional statistics published by national statistical offices (NSOs) are made available through the WorldWide Web. Overall there has been an increase in the number of websites of NSOs from 126 in 2003 to 187 in 2005. Besides, more and more NSOs are discovering the potential usefulness of cartographic functionalities on their websites.

7. Data can be retrieved in the form of a preconceived thematic map on half of the websites. A quarter of the sites have a map interface to find and select regional statistics. Almost one quarter of the websites allow online access to a database to prepare individualized output. And finally 19 websites in total allow individualized output in the form of a thematic map and/or interactive cartographic visualization.

**Diagram 1**

One of the regional interfaces to national statistical office websites

![Diagram 1](http://kartoweb.itc.nl/nso)

**B. The role of maps at Statistics Netherlands**

8. Ever since digital and interactive maps appeared on the internet, Statistics Netherlands has been working on cartographic dissemination and visualisation of regional statistics. Examples of applications are: *Buurt in Beeld* (‘Neighbourhood in the picture’) in 2001 and the *Cartographic Entry into StatLine* (2003). In a 2005 study carried out by the International Institute for Geo-Information Science (ITC), Statistics Netherlands turned out to be among the pioneers within the world of the NSOs. In 2003 only six websites, worldwide, offered users a geographical interface for finding and retrieving data. What is
more both Statistics Netherlands’ applications can be used as a means of presentation and for online analysis and exploration.

9. The Cartographic Entry was introduced in 2003. It is not an isolated application, but forms an integrated part of the online statistical database StatLine. It contains a helpful cartographic interface for finding and retrieving regional statistical data from the level of municipality up to groups of provinces. Between the two Neighbourhood in the picture was the real pioneer. Introduced in 2001 as an isolated tool on the website, it became an instant hit. It provides users with information up to district and neighbourhood level. Information that, up to that day, had not been available through the website.

Diagram 2

10. A clear disadvantage was that users had to install a plug-in for both applications. Nevertheless servers were crashing due to the large amounts of visitors. Both the numbers and the growth rate of visitors were spectacular.

C. Usability and technology: the next step?

11. Notwithstanding the popularity, at a certain moment, the question was raised whether all this cartographic functionality actually helped the users, in particular those of the StatLine database. A user research project was executed. It proved that a cartographic interface was indeed useful for the retrieval of regional statistics. User input also lead to the belief that the potential was not yet fully employed.

12. Meanwhile the Worldwide Web was changing rapidly. Statistics Netherlands’ website was modernised into a professional content management system (CMS). Investing in the - by then old - technology of the neighbourhood application was not considered feasible. After half a decade the pioneer was outdated and screeched to a halt. Several questions emerged during a period of reflection. A new generation of neighbourhood dissemination was desirable, but how should it look?

13. Firstly there is the issue of data availability at the neighbourhood level. The amount of available information is reversely proportional to the number of regions. At the national and provincial level data availability is elaborate. Reliable figures at the neighbourhood level however, can only be computed on the basis of large samples or integral sources. Deepening the Cartographic Entry with districts and neighbourhoods would clearly result in data inconsistency between the regional and the local level. Secondly, user research also made clear that integration of neighbourhood cartography into the still existing Cartographic Entry was not a good idea. The existing user interface was inadequate for
finding neighbourhoods and their statistics: too many actions were required before establishing the demanded outcome. Therefore once again a separately operating application had to be developed.

14. Technology-wise the Cartographic Entry, as part of the StatLine database, was made possible by the application of server-side generated SVG (Scalable Vector Graphics), and from 2010 onwards by the use of Flash. Nevertheless applying the cartographic functionality of SVG or Flash to the neighbourhood level would lead to a great deal of technical problems. Displaying 11 thousand neighbourhoods in addition to the five hundred municipalities would probably result in a user-unfriendly instrument. Furthermore by now Google had entered the scene with some promising new opportunities in the field of interactive mapping.

15. Finally, the arrival of several new and important data sources in combination with Google lead to the belief that a next step in disseminating neighbourhood statistics was viable. Sections 3 and 4 will deal these new perspectives.

III. The use of registers: opportunities and challenges

16. Regional statistics focus on bringing together and combining figures on different subjects and different areas and thus generate additional value. Both the figures and the areas are elements that have gone through enormous changes.

A. Use of registers: data (figures)

17. In 2003 and 2004 the law governing Statistics Netherlands was reviewed and revised. One of the principle changes was that Statistics Netherlands could make use of register-based information, explicitly and free of charge, from other public sources. At the same time the demand for information on the local level was increasing. This lead to a shift in resources which meant starting a separate neighbourhood statistics project within the larger innovation framework called ‘Spatial, housing and mobility statistics’.

18. In retrospect the results are impressive. Figures at the neighbourhood level had been rather classic, with data on population, housing, income and surface area. In 2001 the complete palette contained only forty figures. The use of new registers led to information on new topics: education, employment, social security, car possession and land use. The number of figures available at the neighbourhood level increased to well over one hundred. Work is still in progress. Currently a theme called nearness of amenities is being developed of which the first results were published in 2009.

19. In addition, the two-year publication cycle of the neighbourhood statistics was abandoned. After 1995, 1997, 1999 and 2001 a new publication was constructed with the year 2003 as its point of departure. From then on figures were published on an annual basis. Finally the moment of publishing was speeded up. The 2001 figures were released after 25 months. It took eleven months to publish the first 2005 figures, more than twice as fast. The aim is to publish within nine months, almost three times faster than the old cycle.

20. Although the integration and the extension with the above mentioned themes resulted in numerous inconsistency and disclosure issues, these past few years have lead to a threefold improvement in data-availability: more, more frequent, and faster.
B. Use of registers: geography (areas)

21. Two important geographical registers became accessible for statistical use because of the new law governing Statistics Netherlands. These registers contain information on roads, from highways to streets, and data on address-location, that is x and y coordinates. In the middle of 2004 this new register-information was confronted with our own existing geographical information for the first time.

22. Neighbourhood statistics up to 2004 were calculated using two important geographical sources, the Geographical Base File (GBF) and CBS/TOP boundaries. The GBF is an address register that contains information on residence, street name and house number for all seven million addresses in the Netherlands. It had been kept up to date for decades, but it was common knowledge that it contained faults originating from the non-digital era. The CBS/TOP boundaries reflect the boundaries of all neighbourhoods in the Netherlands. As well as all the polygons (boundaries), this register incorporates the names and codes of all 11 thousand neighbourhoods. Combining these two registers with a third, such as the population register, generates the outcome of population figures for the various neighbourhoods.

23. With the National Road File (NRF) and the Address Coordinates Netherlands (ACN) the missing link could be added: location. Combining all registers gave two results: addresses in GBF that were in the wrong neighbourhood (had the wrong neighbourhood code) or polygons in CBS/TOP-boundaries that were drawn in inaccurately. With the combination of roads and address locations, supplemented with aerial photos that were also available, the quality of our own sources and as a consequence of the neighbourhood statistics could be improved. Two examples demonstrate both situations.

Diagram 3
Quality improvement in Bleiswijk
24. Finally the neighbourhood was changed for over 50 thousand addresses, approximately one percent of all addresses. And in total around one hundred boundaries were altered as well, also around one percent. Overall not the most impressive percentages, but, as expected, the majority of addresses were clustered.

25. From 2004 on both geographical registers are being used in the updating process. The yearly update of particularly new housing addresses used to be a time-consuming and costly activity. The use of NRF and ACN meant the end of municipal surveys and sending out field workers for determining neighbourhood codes for new addresses. A big saving for the government as a whole and Statistics Netherlands in particular. With all this quality improvement is more than just a bonus.

C. **Improvements and inspiration: the next step!**

26. So far we have seen progress in data availability, quality and cost reduction. Data availability improved in quantity, frequency and speed and with a better quality of geography the challenge is now to turn it over to the user. In the past the geographical component was rather invisible. Improving the quality of geography however has lead to new inspiration for disseminating neighbourhood statistics. And with Google entering the scene it is time for a next step in geo-visualisation.

IV. **Statistics and Cartography: recent developments**

27. After making an inventory of user input and after the substantial quality improvements, it was time to work up to a new generation of neighbourhood statistics dissemination. But what should it look like?
28. In recent years there has been a tremendous increase in the use of map-based applications for common tasks, such as checking traffic jams or local weather forecasts. Purchasing houses is another example, as are the travel agencies as early adopters. People are more and more familiar with mapping engines for everyday tasks. 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.

29. Moreover there has been a tremendous increase in the use of streaming photo-based mapping technologies in the Netherlands. This is advanced technology for smoothly displaying maps on the users’ screen and will be explained further on. One cause of their success is probably that the displayed images of the Netherlands 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 why these technologies are such a great instrument for disseminating neighbourhood statistics.

A. Failures and success: a three horse race

30. The year 2006 marked the first step, but a few months along the way there were doubts. The goals were very ambitious, but there were also too many uncertainties. On top of that the expectations of the stake holders were shifting and the project suffered under multiple technical and software restraints. Regrettably there was no successful outcome, only a lot of lessons to be learned.

1. The first success: initiative number one

31. In the last stages of 2006 Statistics Netherlands was asked to join a public-wide initiative within the framework of the national geodata infrastructure, called OnzeGeo, (translated: OurGeo). Being a sizable project, the disadvantage was that we did not have a whole lot of influence on implementation and direction of the project. Internally we were still going over the earlier show stoppers. The government project however was easily accessible and the decision was made to participate and learn. It gave us the opportunity to reflect on the starting points, while at the same time “a first horse was actually entering the race”. There were great expectations of this first horse and results were anticipated soon. After a mere three months the first results were presented in a beta-version during a national conference. Goal, mission and supervision, however, were and stayed insufficiently clear. Above all, the lack of ownership within the framework led to a provisional end. Nevertheless Statistics Netherlands now had a better notion of world players like Google Earth and Microsoft Virtual Earth that had entered the scene. By now there were better prospects on how to use these popular and promising, advanced cartographical interfaces. Horse number one was out of the race though.

2. Real success: initiatives number two and three

32. January 2007 marked the turning point. Statistics Netherlands took the approach to use modern streaming photo-based mapping tools for the presentation of its neighbourhood statistics. There were only a few alternatives with respect to the technology to be used. Microsoft had Microsoft Virtual Earth, Google has its Google Map technology and the more advanced Google Earth. Open source initiatives called Open Streetmap and NASA World Wind were also available. Experiments showed that Google services were technically most suitable for our purposes. But more importantly, looking at our intended users, the Dutch public, we noticed that the two Google technologies were far more popular than others.
33. Initiative number two had its pillars in the earlier 2006 project that came to a standstill. The project was contracted out and ultimately a dedicated website, called *Statistics Netherlands in Your Neighbourhood* was launched for the public in February 2008. It has been a great success ever since. In the first days more than 20 thousand unique visitors were counted. Right now there are about 200 visits per day. The website uses Google Maps to project statistical variables on a Google Map.

Diagram 5

**Website Statistics Netherlands in Your Neighbourhood**

34. The website does not contain all available figures about neighbourhoods, but only the main and most recent figures on population, households, labour, income and housing. Links to related information are added at the bottom of the screen, such as existing municipal publications in PDF format, the latest articles Statistics Netherlands has published on neighbourhoods, a direct link to Statistics Netherlands’ database StatLine and the possibility for users to provide feedback.

35. During the summer of 2007 a more or less coincidental step was taken. After consulting with Google Netherlands for the second initiative, Statistics Netherlands found out that a statistical layer could be developed on Google Earth with the right know-how. And, not unimportant, more easily and a lot quicker than the new website that was under construction. A third horse had entered the race. By the end of 2007 the result was ready for the public and published as part of the main website of Statistics Netherlands. The latter is a supplementary approach and in effect results from *Statistics Netherlands in Your Neighbourhood*. 
The representation of names and boundaries of neighbourhoods becomes visible if the user zooms in to a specific region. The one hundred year old Statistics Netherlands (CBS) icon appears in each neighbourhood when zoomed in sufficiently. When clicked a standard Google Earth bookmark appears as a balloon on the underlying photo.

At the moment both Statistics Netherlands in Your Neighbourhood and Neighbourhood Statistics in Google Earth are positioned as a portal for StatLine, where the complete dataset on neighbourhoods is published. Both initiatives therefore only contain a subset of the total. The focus is on presenting neighbourhood statistics using tools and media people use in everyday life. The three horse race has ended with two horses over the finish line.

B. How it works: flying, tiling, streaming and some features

Both Statistics Netherlands in Your Neighbourhood and Neighbourhood Statistics in Google Earth are supported by Google technology. Satellite photos or aerial photos are used to display areas 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 user’s computer. This type of map generation is called “tiling” or streaming photo-based mapping technology.

In Neighbourhood Statistics in Google Earth we exploit the streaming facilities that Google Earth offers. The layer for Google Earth contains twenty variables for 11 thousand neighbourhoods and their detailed borders. A naïve implementation would result in a Google Earth layer of 60 MB. A layer of that size is too large to be easily downloadable for the general public. Furthermore it would result in performance degradation. We therefore split the 11 thousand neighbourhoods in 450 chunks, namely the municipalities. A chunk (i.e. the neighbourhoods of a municipality) is only loaded when the user zooms in close enough to a municipality. Our experiments show that if a municipality occupies more than 450 pixels that its neighbourhoods can be loaded into the Google Earth layer. Loading the data does not mean that the neighbourhoods are shown immediately. For many regions this would result in a cluttered display where borders and names are drawn on top of each other. We therefore show a neighbourhood only when it takes more than 150 pixels on screen. This mechanism gives users the feeling of immediate interaction.

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 by Google. The format has been actively further developed by Google and is XML-based. Since 2007 KML 2.2 is an open standard adopted by the OGC (Open Geospatial Consortium). Other GeoBrowsers such as Microsoft Virtual Earth increasingly support KML. It is therefore possible to display the same KML file on different GeoBrowsers. The KML format contains many facilities for displaying markers, info windows, lines, rectangles and even full 3D models (COLLADA).

41. Google Maps contains satellite and aerial images. Google Earth takes it one step further by offering a range of interactive possibilities. The latter does not run in an internet browser but is a downloadable application. The earth’s surface is represented using satellite and aerial imagery portrayed on a three dimensional globe. An efficient “pan and zoom” function allows users to explore any area of the globe. Layers can include height information (three dimensional), offering an extra dimension. Google Earth has limited analytic functions. In *Statistics Netherlands in Your Neighbourhood* more attention is paid to this, for instance by including graphs and tables.

42. Google Earth was launched in 2005. A Dutch version has been available since May 2007. After the introduction a worldwide search for spectacular and controversial images started. Three arbitrary examples out of the infinite supply.

Diagram 7
**Three detailed images from Google Earth in the Netherlands**

43. Since Google Earth was launched some of the images have resulted in widespread media attention. Also the parts that were classified, visible as blurred images, generated a lot of attention. In various countries there was a public debate about sensitive areas such as nuclear power stations, military objects, space agencies and royal accommodations. New usages of Google Earth persist. New applications appear in the media almost on a weekly basis. And Google, the parent company, of course is active in engaging new partnerships.

**C. Goals and opportunities**

44. One of the main goals of making a next step in disseminating neighbourhood statistics was to present them in a more natural way and in a context in which they are already familiar. The combination of the familiarity with Google and the recognisability of their own neighbourhood makes Google technology a powerful and excellent way for disseminating Statistics Netherlands’ neighbourhood statistics. What’s more, the local character of the neighbourhood information is now available in a more advanced way than was possible in classic search engines.

45. Both initiatives are supposed to generate ‘traffic’ on Statistics Netherlands’ website. And so they have! Various references have turned up in national and local media as well as in specialist magazines of some branches of industry. The recognisability of the neighbourhood is also pulling new target groups to Statistics Netherlands, and with that existing publications are placed in the shop window. More importantly it catches the eyes of both existing and potential respondents. The neighbourhood statistics in their present form are an investment in the future. They reduce the distance between our National
Statistical Institute and the public. A whole world has opened up, especially for the non-professional user. Statistics Netherlands has developed from an institute that produced traditional statistical tables and publications into a frontrunner in innovative ways of data dissemination.

V. Looking forward: future perspectives

46. The use and also the production of maps has democratised with the internet. The results often look good, certainly from a technical point of view. They can, however, present an image of reality that is completely wrong. Producing an accurate (thematic) map is far from obvious. The effects of choices in symbols, colours, generalisation of boundaries and the scale used are often overlooked. Some users are genuinely under the impression that the images in Google Earth are actually real time. With that we are back at Carl and Nyls. And from this we can conclude that as a national statistical bureau we carry a great responsibility in disseminating statistical outcomes.

A. Solving the dilemma

47. Maps are a simplification of reality. Thematic maps have the design rule that only the geographical distribution of the data should be shown. All irrelevant details that distract the user should be left out. The use of aerial photos means reintroducing reality into the map. For the neighbourhood statistics aerial photos have an advantage: users recognise their house, street and neighbourhood by using the zoom option of Google Earth. For creating a thematic neighbourhood layer we have to consider the effect of combining abstract maps with the realistic images in Google Maps and Google Earth.

48. The dilemma is how to combine a thematic map where geographical features are irrelevant with the very detailed aerial imagery that Google Earth offers. We found one solution to this problem.

49. In our solution we mix a thematic map with aerial photos by using transparency. Furthermore we make the transparency dependent on the zoom level. When a user starts with the country overview, the thematic layer (choropleth) is opaque: it is just a thematic map juxtaposed on Google Earth. When a user starts zooming in we increase the transparency of the layer, and effectively mix the thematic map with the aerial photos. In this solution users smoothly shift from a pure thematic layer to a mixed version where they can recognise their neighbourhood. In the experiments we conducted we generated chloropleth maps that are very detailed (using the neighbourhood polygons). At country level these maps show detailed neighbourhood patterns but no aerial photos. When a user zooms in on his/her neighbourhood the aerial background starts to shine through. An example of a subject that can be represented in this way is the population density in neighbourhoods.
B. Future work

50. This paper described the steps Statistics Netherlands has taken to get to the next step in disseminating neighbourhood statistics. Instead of using traditional mapping techniques, it has introduced streaming photo-based mapping techniques such as Google Maps and Google Earth. These current initiatives are very powerful in helping users 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 feature for a district, municipality or the Netherlands as a whole. This would require a more advanced mechanism possibly using heat maps to display the status of the Netherlands at a glance. Google Earth offers captivating possibilities for the use of both colour and 3D models in presenting statistical variables. Early experiments with these facilities are very promising.

51. Another promising area is the addition of animated statistics to the Google Earth layer. Google Earth also incorporated a facility to add time to layers, thus making it possible to visualise statistical trends in the context of streaming photo services. This can be used to show changing neighbourhood boundaries or changes in municipal population over time. Early experiments with this facility are also promising.

52. A drawback in the present design of both initiatives is the dependency on Google. It is common knowledge that Google registers what users of its services do on the internet. And there are other disadvantages. Alternatives for Google are being developed within the
government. As soon as a commercial free alternative is available Statistics Netherlands will redetermine its position in the field of photo-based mapping technology.

53. Driven by European initiative INSPIRE, Statistics Netherlands also uses agreed standards for publication of its geo information. In the near future even more advanced technology such as Web Map Services (WMS) and Web Feature Services (WFS) will be used and available for other portals.

54. The most promising development finally, is further exploiting the use of register information. The combination of GIS and administrative sources offers a new range of options. In 2009 a so-called ‘backbone of addresses’ was completed and implemented which enables new spatial statistics. This will eventually result in spatial outcomes in the field of sound barriers, water floods, or the concentration of fine particles in the air. Outcomes that are no longer necessarily based on administrative boundaries. This will open the way to even more interaction, also in geography. In conclusion: we are only just beginning.

VI. References


Beeckman, D.J.D. (2003), Kaarten en kaartgebruik bij het CBS. CBS website.


Beeckman, D.J.D., de Jonge, E. (2008), CartoGraphy with a capital G: Google and more. CBS website


Verweij, P. (2003), Geografische Informatie Systemen in de journalistieke praktijk.

Web-visualisations (some interesting examples):

- FlowingData, Examples of web-visualisations: http://flowingdata.com/
- GENeSIS (a research node of the National Centre for e-Social Science in the United Kingdom), Examples of web-visualisations: http://www.casa.ucl.ac.uk/genesisblog/?cat=18.