

**UNITED NATIONS  
ECONOMIC COMMISSION FOR EUROPE**

**CONFERENCE OF EUROPEAN STATISTICIANS**

**Seminar on New Frontiers for Statistical Data Collection**  
(Geneva, Switzerland, 31 October-2 November 2012)

Topic (ii): New methods and technologies

**INNOVATION AT STATISTICS NETHERLANDS**

**Contributed Paper**

Prepared by Barteld Braaksma, Nico Heerschap, Marko Roos and Marleen Verbruggen,  
Statistics Netherlands

**I. Introduction**

1. We notice several trends in society that ask for an appropriate response from statistical offices. A few examples: the political pressure to reduce the administrative burden, respondents (both households and companies) who are less willing to respond to time consuming questionnaires, the high volatility of information and the increasing need for rapid, to-the-point and easily accessible information, the shift to mobile devices and finally, the increasing importance of internet.
2. It is the challenge of statistical offices to use their knowledge and innovation power to the optimal extent in order to remain able to respond pro-actively and in a creative way to these developments. This paper describes how Statistics Netherlands has developed an Innovation programme. A three-stage funnel approach plays a key role in this programme. This approach gives maximum room for bottom-up development of ideas, while its focuses at the same time on maximum contribution to the goals of the organization. The Innovation Lab is an important instrument for the Innovation programme. It offers a suitable environment to support the generation of ideas and test their feasibility.
3. Statistics Netherlands has identified four priority areas for innovation. Among them is innovation in the area of data collection. We will describe two examples of recent innovative ideas that are adopted by the Innovation Lab: Internet as a data source and data collection using smart phones.

**II. The Road to Innovation**

4. In January 2012, Statistics Netherlands formally started its Innovation programme in order to accelerate innovation and thus facilitate dealing with the trends signalled in the Introduction. Both an approach for channelling the stream of ideas and infrastructural facilities for experiments have been established.
5. Before the programme actually started, a number of study visits were paid to government institutions and private companies renowned for their innovative character in order to see how they deal with innovation. This gave a lot of inspiration for our own Innovation programme.

Many best practices from other institutions could be copied and incorporated into our own Innovation programme. This gave the programme a head start, saved the investment of inventing an own approach and increased the chances of success.

## **A. Three Drivers**

6. The Innovation programme is driven from three sides:
  - (a) External developments. New output channels are emerging that require attention. One example is the move towards Open Data. All Dutch government institutes are required to make their public information available in open data format. Since information production is the core activity of Statistics Netherlands, we are expected to play a leading role in this movement. Another challenge is the reduction of administrative (response) burden in particular for companies;
  - (b) Technological challenges. Increasingly, many new so-called ‘big data’ sources are becoming available, which need new processing techniques and also provide new output opportunities. Examples include data from social media like Twitter, mobile positioning data, prices collected from internet and traffic detector loops (over 10,000 measure points built in highways in the Netherlands to measure traffic intensities on a minute-to-minute basis);
  - (c) Internal ideas. Many employees at Statistics Netherlands have ideas to improve existing statistical processes, create new output based on the re-use of existing material, or generate synergies across different subject matter domains. A proper outlet for this creativity was lacking. Many ideas could therefore not sufficiently be explored, to the disappointment of their inventors.
7. All three drivers generate innovative ideas, where an idea is to be understood in the broadest sense possible. These ideas can be explored and tested using the approach sketched below.

## **B. Three-Stage Approach**

8. *‘The best way of having a good idea is having a lot of ideas’*. A three-stage funnel approach was introduced in order to manage the stream of ideas and select the most fruitful and promising ideas. The funnel approach was inspired by a model that is successfully implemented by other organizations.
9. The first stage of an innovation track consists of “Idea Generation”. People having a rough idea are stimulated to submit it to the Innovation programme, which then helps to enrich the idea into something that appeals to a possible sponsor. Enrichment may mean either better articulating the idea, improving its focus or combining similar ideas. When the idea is sufficiently clear, a sponsor is sought who is interested in the idea and has the intention to implement the idea when it is proven successful. The sponsor is often a manager of a production unit.
10. The second stage is the “Proof of Concept”. When a sponsor has been found and the idea is sufficiently clear, concrete and focused, a proposal for a Proof of Concept (PoC) is formulated. The PoC is a well-defined experiment that can be conducted in a limited time (three months as a rule of thumb) and uses limited resources (500 working hours at most). Both the sponsor and the inventor of the idea are expected to contribute resources to realise the PoC. The Innovation programme has a supporting and facilitating role.
11. Efforts are made to keep the overhead and paperwork, associated with running an innovation track, as light as possible. For example, the PoC is formulated using a one-page template to describe its main characteristics and it does not require an elaborate project plan that has to be approved by a steering committee. The Innovation programme also has a pre-set budget for methodological and/or IT resources, which are often needed for PoCs. The PoC is expected to deliver a short report explaining its approach and results.

12. The third stage is that of “Implementation”. It is mainly the responsibility of the sponsor to decide on implementation, based on the results of the PoC. This may for example mean that a regular development project is started. The innovation track can also stop after a PoC, if the results are considered not fit for implementation. Other outcomes may be to start a further PoC or start a research project to improve the theoretical foundation of a PoC’s results.
13. The funnel aspect means that our aim is that about half of the ideas are developed into a PoC, and that roughly half of the PoCs are followed up by an implementation project (or other follow-up activity). This differs from the traditional approach for research and development projects, where basically all projects started are expected to deliver (applicable) results. It also implies that we have to accept the risk that an idea will not lead to implementation and learn not to consider this as a failure. Exploring an idea in itself is worthwhile, even though it is not further pursued.
14. The approach described here serves as a guideline and is not adhered to very strictly. If another approach seems more suitable to follow up on an idea, we will choose this other approach. This is in line with the philosophy of the Innovation programme, which tries to facilitate and enable ideas without too much hassle. The one thing we adhere to strictly, however, is that we only follow up ideas for which a sponsor has been found. If nobody is willing to consider using the idea, we do not want to invest in it.
15. Many innovation tracks are done in close collaboration with external partners. These include non-profit institutions like universities and government bodies and commercial enterprises like IT consultants and data providers. In this way we gain access to knowledge, expertise and data sources that are not (yet) available in our own institute. Often a win-win situation arises. Some partners for example appreciate the fact that they can test or validate their products and services on real data. Others are interested in the statistical results for their own purposes.

### **C. Research vs. innovation**

16. When the Innovation programme started, one of the recurring questions was what distinguishes innovation from (methodological) research. This is indeed a non-trivial question, and in some cases there is overlap between research and innovation. Moreover, often the same people work on both research and innovation activities.
17. The main characteristic of (methodological) research is that it starts from a theoretical perspective and tries to solve an abstract problem. This is often not straightforward and the results are uncertain. The abstract problem may or may not be inspired by a specific situation. Usually, only after the abstract problem is solved, one tries to find practical applications and possibly implementation through a development project.
18. The main characteristic of innovation is that it starts from a concrete, practical problem, including an idea how to solve it. A Proof of Concept serves as an experiment to demonstrate that the proposed solution actually works. Of course, when carrying out a PoC many smaller and bigger (methodological) issues may arise which need to be addressed along the way.
19. Given the above distinction, both methodological research and innovation have their own niche and are valuable in their own right. There are many opportunities for cross-fertilization. For example, the practical applicability of a research project may be demonstrated by a PoC. On the other hand, the outcome of a PoC may be that more in-depth research is needed for what at first looked like a straightforward idea with a clear-cut solution. Moreover, a PoC may use existing methodology or the results of a PoC may suggest methodological research topics.
20. A Board for Research, Innovation and Standardisation has been established to monitor relations and interactions between innovation and research on an on-going basis. It is chaired by the Director for Methodology and Statistical Policy and includes senior management representatives.

## **D. Organisation and governance**

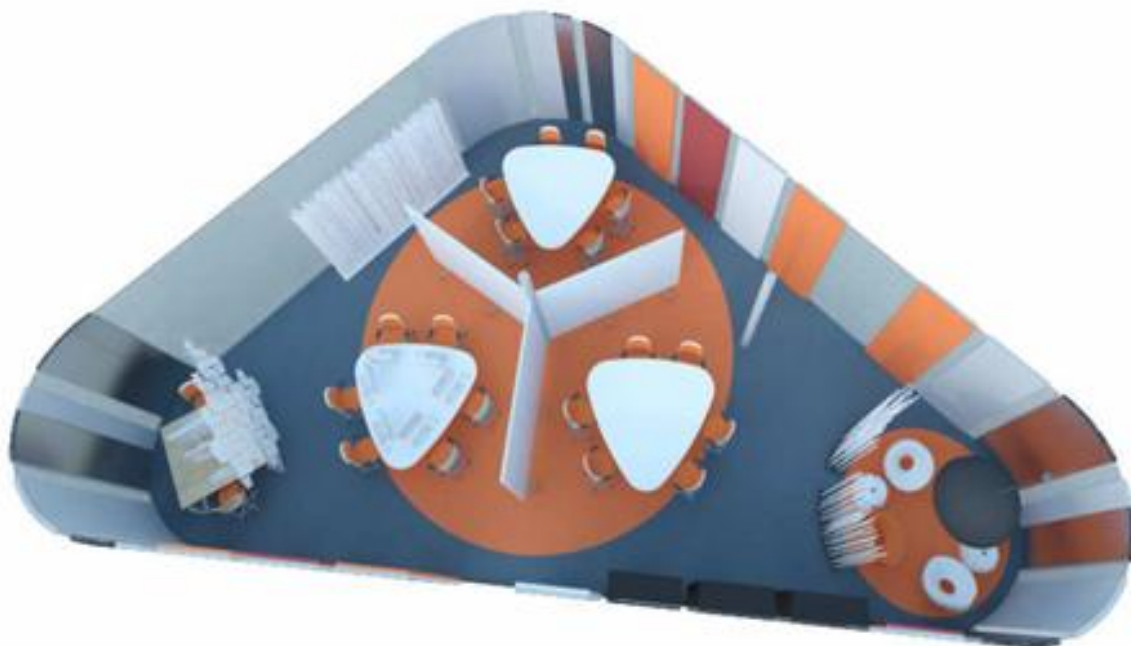
21. The Innovation programme is essentially run by three people: the Programme Director for Knowledge and Innovation, the Innovation Programme Coordinator and a Programme Assistant. All three people devote about half of their working hours to the Innovation programme. Their main tasks are to make sure the Innovation programme and facilities run smoothly and to support the innovators where needed.
22. The actual innovation work is done in the various activities like PoCs. Usually the original inventors of an idea work on their innovation tracks, together with representative from their sponsors and, when needed, people from the methodology and/or IT departments. A pre-set annual budget is available from which methodology and IT resources may be drawn.
23. The Innovation programme falls under the direct responsibility of the Director General of Statistics Netherlands. It follows the normal Planning and Control cycle of Statistics Netherlands, which means that an annual work plan and quarterly progress reports are submitted to the Board of Directors for discussion before formal approval by the Director General.

## **III. Infrastructure and facilities**

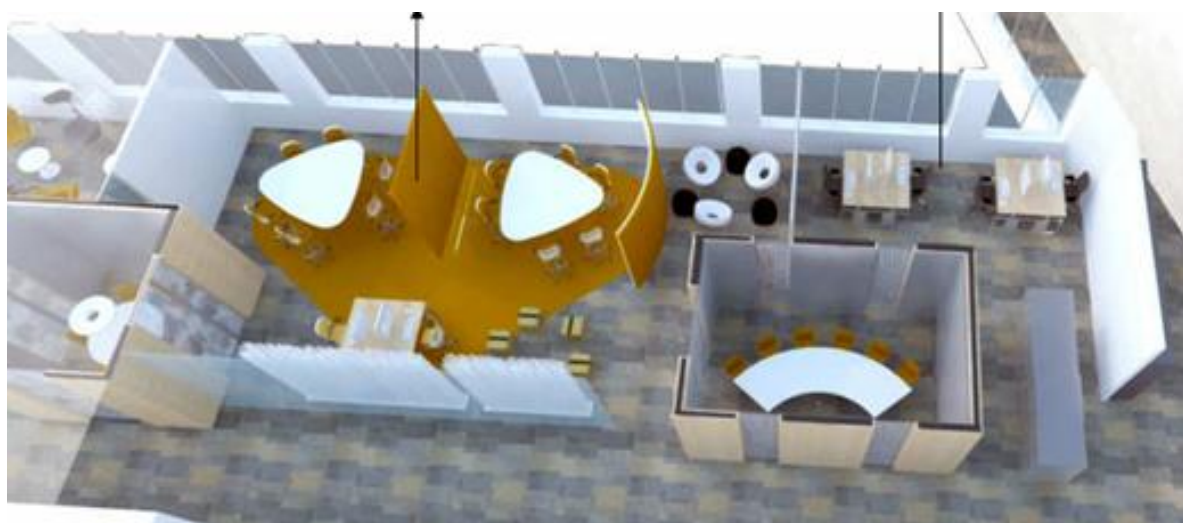
### **A. The Innovation Lab**

24. The Innovation Lab is a physical space (actually two rooms: one in The Hague and one in Heerlen, connected by video conferencing facilities) that encourages quick elaboration and testing of innovative ideas and collaboration. It was officially opened by the Director General and Deputy Director General in May 2012. The Innovation Lab is now being used to work on several innovative ideas. Some examples are given in Chapter IV below.
25. Many ideas need the so called “attic environment” to grow. The Innovation Lab offers this facility. In the Innovation Lab, statisticians can work on laptops and desktop PCs without the restrictions imposed by the closed internal network and its standard workstations. This makes it easier to test new methods, try out non-standard software, and simulate alternative statistical processes.
26. The Innovation Lab can also be used as a collaboration space. Every piece of furniture is on wheels, which makes the use of space very flexible. It can be used either as a single space or divided in up to three separate workspaces. This makes the Innovation Lab a suitable environment for brainstorm sessions, workshops, informal presentations and open coffees etc. The Innovation Lab stimulates collaboration between different fields of expertise. It is an inviting and inspiring environment.

*The Innovation Lab in The Hague*



*The Innovation Lab in Heerlen*



27. The organisation of workshops has rapidly become one of the key uses of the Innovation Lab. Workshops have proven at Statistics Netherlands to be a powerful tool for brainstorming and focus at the start of a project, to reach a breakthrough at a difficult stage of a project or to exchange experiences between different projects / at a specific field of interest. Statistics Netherlands has a pool of facilitators that have been trained to organise and facilitate workshops. They are very experienced in: formulating the aim of a workshop, the issues to be solved, designing the workshop programme and work methods, and chairing the meeting. The Innovation programme works in close collaboration with these facilitators e.g. by the organisation of workshops around the special priority themes of the Innovation programme.

## Workshop Output Innovation



### B. Other infrastructure

28. Apart from the Innovation Lab, the Innovation programme has a number of state-of-the-art stand-alone laptops and desktop PCs available that can be used to carry out experiments that cannot be easily done on the standard closed network environment. This may be either due to the fact that specific software is not available on the network (which allow for obvious reasons only extensively tested and approved software), or that the standard virtualised desktop PCs have limitations in storage, memory or processing power which makes it awkward to use them for experiments with for example big datasets. Since some of the data being processed are sensitive and should be stored securely, safes are available to store for example data DVDs, laptops or hard disks.
29. On the stand-alone laptops software can be downloaded and installed at will. The Innovation programme has a limited budget to buy specialised software for testing purposes. This budget may be applied to buy specialised hardware as well.

## IV. Innovation Tracks

### A. Priority areas

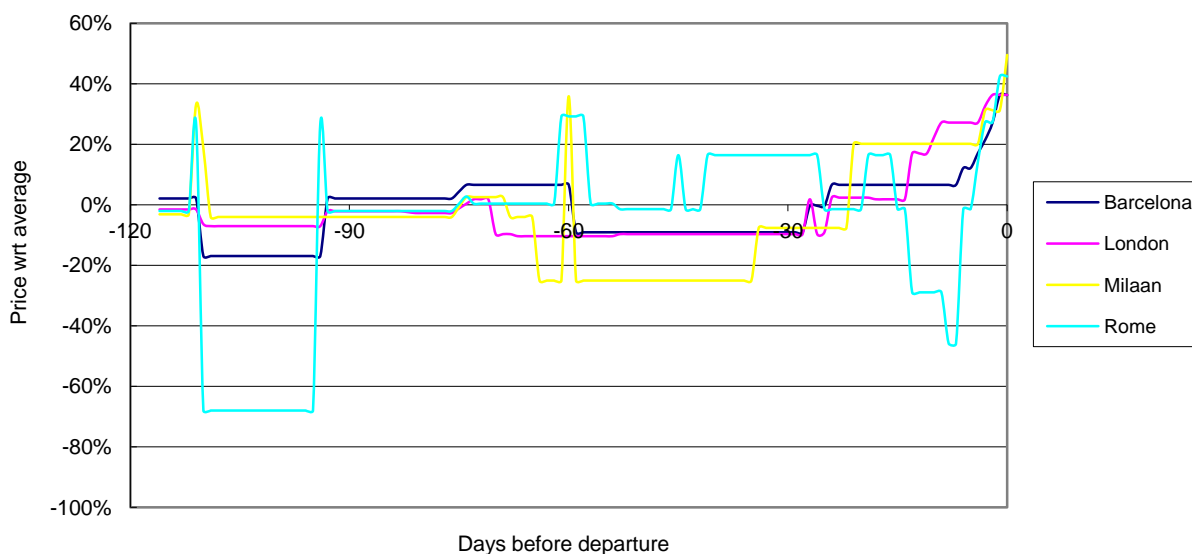
30. The Innovation programme is loosely organised in four priority areas.
  - (a) Data Collection Innovation
  - (b) Efficient Processes
  - (c) Output Innovation
  - (d) Big Data
31. Here, 'loosely' means that we do not simply dismiss ideas that do not fit in either of these four areas. We do, however, stimulate ideas that do fit in these areas and in practice all ideas submitted so far fit in one of them (and some fit in several categories at once). The areas are chosen because they correspond with goals of the strategic multi-annual statistical plan. The same areas are distinguished in the research programme, which is natural given the strong interaction between research and innovation described above.

32. All four areas are rather active. Below we elaborate two innovation examples in the area of Data Collection Innovation.

## **B. Internet as a Data Source**

33. The internet has become an indispensable infrastructure for society. A growing proportion of our day to day communication, information and economic transactions now take place via the internet. The distinction between the physical and virtual world is becoming increasingly blurred. This is reinforced by the introduction and rapidly growing penetration of intelligent phones or ‘smartphones’. These smartphones and other mobile devices supplant the desktop PC as the main access to internet: the mobile revolution.
34. There is a growing need from society and policy makers for reliable, up-to-date and high-frequency statistics on the development of internet itself, as a complex and rapidly changing phenomena. It is difficult to satisfy these needs with traditional data collection methods, because of the decreasing willingness to participate in surveys, limited resources and limitations of the survey method itself (frequency, number and type of questions that can be asked and periodicity). To solve this problem more and more is pointed to internet itself as a new data source for making statistics. This is possible because all activities on the internet leave behind digital footprints, which can, with new technologies, be measured 24 hours per day and 7 days per week. In fact, the enormous and rapidly growing amount of data on the internet not only concerns the developments of internet itself, but contains information about almost “everything”.
35. ‘Internet as a Data Source’ or IaD, first of all, concerns innovative techniques for searching and retrieving information available on the internet. The data collected then can be used to produce statistics. An example is the observation of prices of goods and services using special software (“internet robots” or “crawlers”), with the purpose to calculate inflation figures. However, practice shows that IaD requires much more than just a set of innovative technologies and the collection of data from the internet. In particular, the methodology and processes underlying the technologies need proper attention in order to translate the unstructured data collected from webpages to reliable and representative statistics. This includes for example the categorisation and standardization of collected information, de-duplication, visualization and interpretation of internet data. And, finally, IaD has harsh legal dimensions, which have to be met before an IaD-project can be started.
36. Mid 2010 a two year IaD research programme was set up within Statistics Netherlands. This programme included research projects on internet robots and on the use of datasets, that were directly obtained from website owners. Moreover, research projects on the use of smartphones (see next paragraph) and on the use of new technologies in the transportation sector (GPS, RFIDs, Bluetooth, ...) were also included. The programme included both theoretical research and experiments. An important goal was also to investigate opportunities for incorporating IaD elements in production processes. When the Innovation programme started, it adopted the Proof of Concept (PoC) part of the on-going IaD program. The Innovation Lab and its infrastructure supplied the necessary IT facilities for processing the large data sets involved and enabled the use of non-standard software for building internet robots, among other things.
37. Statistics Netherlands began experimenting with internet robots for collecting price information. The first experiment concerned the observation of prices of airline tickets. One of the conclusions was that, due to rapid and unpredictable changes in website layouts, it was not feasible to build a stable internet robot for production purposes. However, the experiment did give insight in peculiarities of ticket prices like their day-to-day fluctuations (see figure below). Experiments with price robots followed in other domains such as petrol, clothing and cinemas. Moreover, an automation tool was built to support the manual collection of internet data for the calculation of price indices.

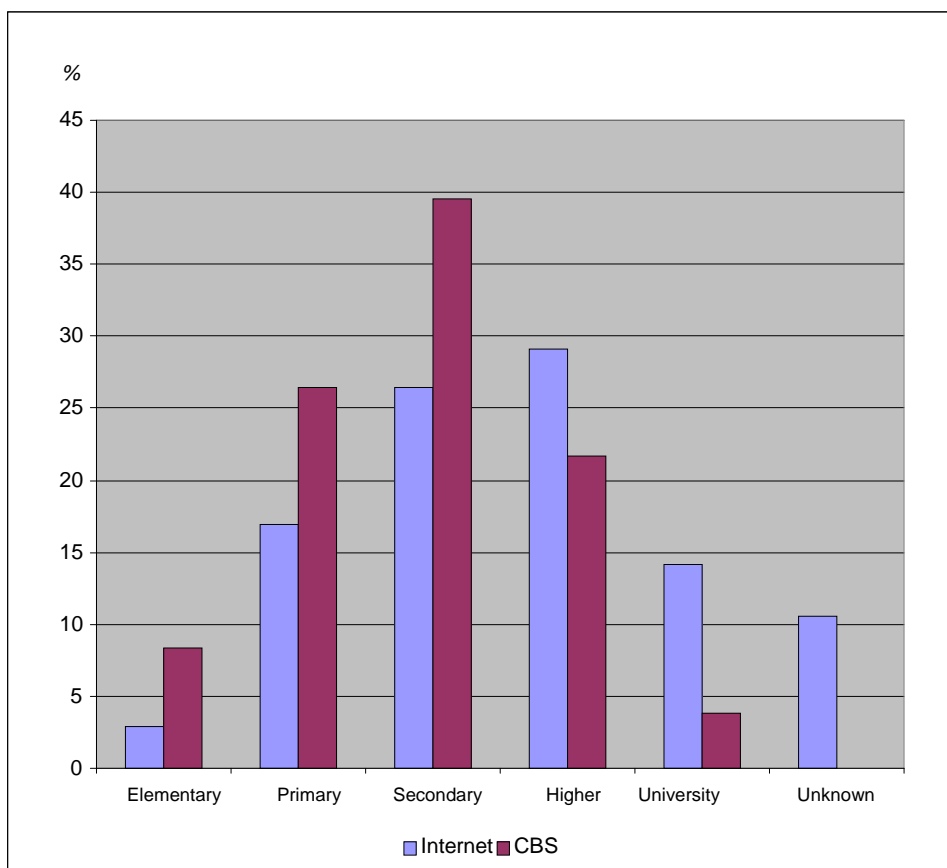
*Day-to-day fluctuations, obtained from internet robots, of airline ticket prices to four destinations*



38. Internet robots have also been used to collect data on job vacancies. The internet data were compared with our quarterly statistics on job vacancies. So far, the results of this project are somewhat disappointing. The internet data appear to be biased, in a way which is (with current knowledge) difficult to correct, and the financial business case for implementation is negative. However, research is still carried out to find other useful applications of these internet data.
39. A third area where internet robots were used is that of housing. Information from the internet is retrieved on houses for sale and for rent. Compared to the project with job vacancies, this is a rather successful project. The main reason is that, in contrast to the job vacancies project, the internet data on housing can directly be related to the central framework of all houses in the Netherlands. This also provides the possibility to add already available data to the dataset and conduct more in depth analyses, for example on the changing tension in housing market.



## Job vacancies from internet robots and quarterly statistics compared



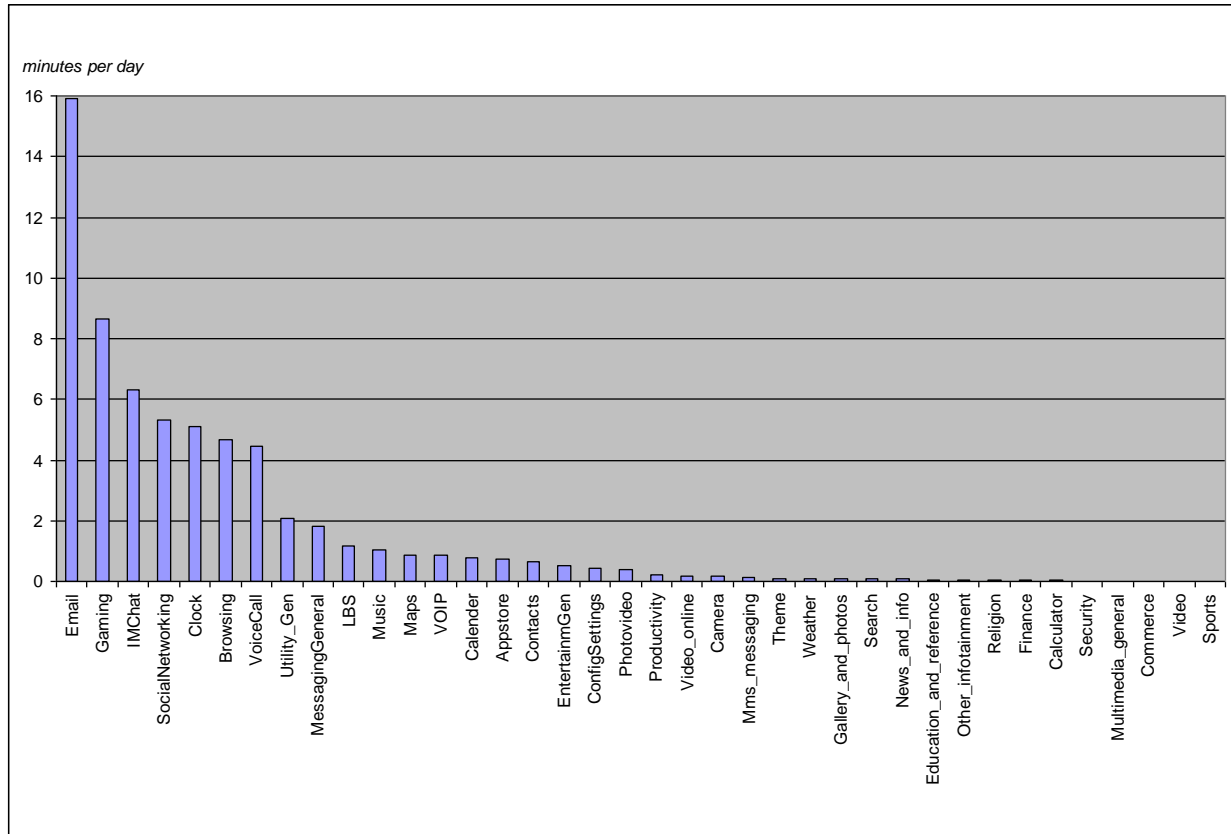
40. An example of the use of data from a website owner is that of the data from the Dutch EBay for the years 2006 – 2011. These data on the consumer to consumer market are analysed in the broader framework of the internet economy. Another example is the use of data from an internet speed measuring site. It offers internet users the opportunity to analyse their actual broadband speed and, simultaneously, collects data on background characteristics, location and provider subscription. This provided possibilities to analyse relations which are very useful for policy reasons.

### C. Data Collection using Smart Phones

41. Another area where data collection can be innovated is the use of smartphones and other mobile devices. More and more people are becoming owners of such intelligent mobile telephones. It is estimated that one out of three persons in the Netherlands already owns a smartphone, and this is rapidly growing. It must be said that smartphone studies are not only conducted because of the possibilities to innovate the data collection process, equally important is the fact that there is a big need for data which provide insight in the developments and use of mobile services and therewith their social and economic impact.
42. Since 2010, Statistics Netherlands conducted several experimental smartphone-based data collection studies. One study involved tracking about 130 volunteer smartphone owners over a period of four weeks in October 2011. They installed a research App on their smartphone, which produced measurements in the form of data logs for the full observation period. These data logs provided useful insights into aspects such as frequencies and duration of the use of mobile services, visits to websites (surfing behaviour), calling and SMS behaviour, data consumption and traffic and the use of the different internet connections. The figure below provides an example. In addition, the location was recorded every five minutes. Experiments were also conducted with a number of so called pop-up surveys. That is to say, immediately after the use of a social media application two questions were asked about the reason to participate in social

media. In addition to the tracking of the smartphones of the volunteers, two surveys were conducted: one before the tracking period started and one at the end of the study. The combination of smartphone measurements, pop-up surveys and complementary surveys provided a very rich collection of data on behaviour as well as on motives and background information.

*Smartphone measurements: intensity of use per type of mobile service (in minutes per day)*



43. There are still big challenges with these kinds of smartphone studies. The main challenge is the recruitment of participants. It appears that at the moment only 5-10% of potential participants are willing to participate in these kinds of surveys. The main reason for not participating is privacy concerns. Other reasons include no time, not allowed to download the App because it is a company owned mobile, or the App consumes too much power. Despite this, the 2011 study was considered a success and will be repeated in 2012.
44. A second study with smartphones involves data collection for mobility statistics. In this study we cooperated (and still are cooperating) with a not-for-profit organisation that specializes in IT-innovation for governmental organisations. They developed a smart-sensing App that measures trips and provides feedback about the trips. It also automatically deduces the modality of the trip, for instance a trip by car or by foot. The App strongly hinges on a back-end infrastructure that deduces much of the trip information. The algorithms used rely on feedback from users and data from other web services.
45. We tested the App with a small group of volunteers, to determine whether it provides data of sufficient quality and a sufficient user experience. We did encounter some problems in this first test. The problems related to increased battery depletion, the precision of the recorded trips and the deduction of the modality. Our partner indicated that for a second phase these problems will be dealt with. We have recently decided to proceed with the next phase, in which additional questions will be asked from the user. These questions will be built into the App itself. We plan to test this App in the first half of 2013.

## V. Concluding Remarks

46. The Innovation programme has made a successful start. In the first half of 2012 already more than 30 ideas have been submitted to the Innovation programme. This is more than what was expected (25 ideas for the whole year 2012). First experiences with the Innovation Lab have also been very positive. In the first months after the opening almost 20 workshops were organised in these facilities! The technical facilities and informal ambiance are highly appreciated.
47. The Innovation programme has been very active in organising communication events like an Innovation market, thematic lunch meetings, presentations and open coffees. This has helped to stimulate creation of ideas and raised awareness of the goals and results of the Innovation programme. It has also served as a means to stimulate knowledge sharing between practitioners from different fields of expertise.
48. Although the Innovation programme is still in its infancy, we can already see some types of innovation tracks emerging. For example, one type of innovation track explores a new data source and tests its possibilities for new statistical output. Another type tests if a specific software tool is fit for a statistical task or, alternatively, tries to select a suitable tool from a shortlist for a given task. A third type of innovation track turns available statistical information into new output; for example by applying a new visualisation technique or combining different sources.
49. The Innovation programme facilitates inventors to realize their innovative ideas. In doing so, we often help with non-statistical matters, such legal aspects. When exploring a new data source we always need to be aware of possible confidentiality and security issues. And when collaborating with external partners we have to think about contractual negotiations, which is not an area of expertise for most statistical researchers. Another experience is that some PoCs need experimental software or dedicated hardware which is –for good reasons–not available and not allowed on the standard (closed) IT environment. The Innovation programme helps to build up experience with such issues and tries to find appropriate solutions.
50. After about two years of experience with IaD, the main conclusion is that it is not too difficult to make statistics based on internet data only. To make the step from just internet statistics to official, reliable and especially representative statistics, however, can be a huge challenge. It is not so much the technology and data collection itself that poses the major challenges, but the methodology and visualization in the later steps of the statistical process. It helps if the internet data can directly be coupled to existing population frames. The fact that data and expertise sometimes must be bought from third parties can lead to negative financial business cases. Companies are becoming more and more aware of the value of their (internet) data. Despite these challenges we believe that IaD has the potential to be integrated into the statistical processes of Statistics Netherlands and contribute to quick and new or improved statistics with less survey burden at lower costs. The same conclusion applies to the use of smartphones in the data collection process. A key challenge here is to recruit enough people to participate in such studies.