Official Statistics in the era of Ubiquitous connectivity and Pervasive Technologies

Michail Skaliotis[1], Eurostat

Summary

Scenarios about future developments in ICT (information and communication technologies) and the internet in particular are becoming more and more precise, at least as far as the near future direction of technical developments is concerned. Ideas which seemed to be science fiction some time ago, are now taking the form of tangible research projects and it is certain that a number of these ideas will end up in real life applications. Ubiquitous connectivity and context aware technologies (or pervasive technologies) are at the centre of ICT 2.0[2] scenarios. A predominant characteristic of future ICT will be that data and metadata storage and exchange, information processing / computation, censoring, data linking and reporting will become standard, embedded features in thousands of devices (fixed and mobile) used in everyday life. What do these developments mean for Official Statistics?

The internet itself is becoming a rich data source which can reduce remarkably the response burden on individuals and businesses. The European Statistical System (ESS) should respond positively to the challenge of developing an appropriate quality framework for the measurement of internet traffic flows (ITF) for statistical purposes. Actions in this area will require synergies between official statistics and private source data as well as co-ordination within the international statistical community[3].

There is no doubt that official statistics have made significant progress over time in providing a trustworthy evidence-based platform for policy making, as well as a living mirror of the society and economy. Technology however, is advancing and entering into our daily environment at an unprecedented speed and scale. I believe that despite a number of obvious good examples of progress (e.g. automated data capture, data linking from different sources, Bayesian methods, multiple and multi-level modelling, etc), official statistics remain largely at their original version 1.0.

The aim of this paper is to raise awareness amongst official statisticians of the necessity to move towards official statistics 2.0. A profound technology shift is happening now that requires changes to our professional paradigm and our business model. The issues involved in this process are many and complex; however, the most critical ones are not technical but rather social, legal and …above all 'it is a matter of trust in digital life'. I do not think it is only a *challenge* for official statistics to act now, but also a *responsibility* towards the society. It is very encouraging that the current policy context, visionary leadership and collaborative structures in the wider European Statistical System are very propitious for making 'official statistics 2.0' a reality.

The first section of the paper will review current trends (in) and the increasing importance of data mining; it is argued that--because of a wider awareness of the (real or perceived) power of statistical analyses for explaining and / or predicting numerous events in everyday life--the market of ‘official statistics’ will soon be invaded or replaced (at least partly) by private actors. Then I try to highlight a few possible technological developments in the form of scenarios about the future of the internet and ubiquitous connectivity. These scenarios are not based on my intuition or imagination but rather on current research projects and the opinions of leading technology experts. This section argues that several of what may look as ‘science fiction’ applications will be realised in the near future and that common characteristics of such applications are data analysis, privacy, ethics, trust and ‘information

[1] The views expressed in this paper are those of the author and do not necessarily reflect the official views of the European Commission.

[2] The next generation of ICT in a ubiquitous context: information and communication will be enabled (and embedded by default) in numerous devices of everyday life (not only in phones, PCs, and PDAs, but actually in almost everything, everywhere at any time).

[3] OECD has recently initiated a debate on this issue; see for example WPIIS doc. DSTI/ICCP/IIS/RD(2009)2, Paris, 21.4.2009, 'Building synergies between official statistics and private source data.'

everywhere'. A subsequent part is devoted to a few illustrations of domains in which official statistics could embark with some piloting exercises; we look at the implications of large scale RFID deployment for price statistics, the use of internet as a data source (for deriving / producing statistics) and the future of health statistics. The last section comprises a number of concluding remarks.

1 The power of data and the market for official statistics

We can find numerous examples and scientific papers about the increasing use (and misuse) of statistics and data mining for a wide range of applications both by government and non-government actors. Terms like 'evidence based decision making', 'data driven policy making', or 'statistics and democracy' are very familiar amongst the circles of official statisticians. In my own professional career as a Eurostat official during the last 27 years I have been providing data for the formulation, monitoring and management of several EU actions, which often involved the allocation of substantial community funds. During the same period I have observed an ever increasing demand for statistics, primarily demanded by my fellow colleagues working in the policy departments of the European Commission, but also requested by other (private and public) users of statistics. This increasing pressure for 'more and speedier' statistics is now widely felt upon the entire European Statistical System. As a consequence, the ESS cannot meet all the demands and / or cannot provide the required statistics in a timely manner. A first observation at this stage is that the market of official statistics can potentially accommodate more actors than the NSIs (National Statistical Institutes) and other bodies producing official statistics (central banks, ministries, local authorities, etc).

There is however an obvious question as to the attractiveness or not of this market. In other words, is there any interest for private actors to enter the market of official statistics? The answer is yes. While the interest of private actors at this moment--driven by commercial purposes--is primarily focused in areas which are not the core of official statistics, the increasing awareness of the power of data mining and statistical analyses will result in an important growth of the number of (competent) data analysts who will compete (why not) against NSIs in a number of fields which today belong to the somewhat 'closed' market of official statistics. One of the best books I have read during the last 5 years about the growing importance of statistics is Ian Ayres's[4] bestseller *'Super Crunchers: Why thinking-by-numbers is the new way to be smart'*[5]*.* I am confident that this book can persuade even the most sceptical person that statistical analysis is increasingly becoming an indispensable companion of our decision making process in almost all aspects of our lives. Through a series of lively presented and rigorously thought examples--ranging from disease diagnosis, wine prices, unemployment benefits, education policy, sports, criminal justice to the movie industry--Ayres concludes that *'…any large organization that is not exploiting both regression and randomization is presumptively missing value. Especially in mature industries, where profit margins narrow, firms competing on analytics will increasingly be driven to use both tools to stay ahead….'.* The strong interest of big companies in playing a leading role in the market of statistics is manifested by recent major acquisitions of 'business intelligence firms'. Some spectacular facts are cited by Ayres:[6] *'…Oracle purchased Hyperion Solutions Corp…., SAP purchased Business Objects….and IBM announced the purchase of Cognos. These multibillion-dollar acquisitions of firms whose sole product is number crunching is powerful evidence that data-driven decision making has market value…..'.* If we now think in a context of future internet and ICT developments, the competition in the statistics market will most likely favour those who will have quick access to the relevant digital footprints, i.e. the predominant type of data in the future. Based on current evidence, my deep concern is that the emerging enormous data bases which contain these new types of data, the digital footprints, are owned almost entirely by private actors such as google, internet service providers, social networking sites, e-bay, amazon, super markets and many others. The following examples will highlight this aspect.

A very interesting paper[7] by Hyunyoung Choi and Hal Varian, '*Predicting the Present with Google Trends'*, demonstrates the potential of using search engines' queries for statistical purposes. The authors' motivation is to use Google queries in order to help predict economic activity: *'Economists, investors, and journalists avidly follow monthly government data releases on economic conditions.*

[4] Ian Ayres is an econometrician and lawyer, Professor at Yale Law School and Professor at Yale's School of Management.
[5] Published by Bantam Dell, New York, September 2007, ISBN 978-0-553-38473-4.
[6] Ian Ayres, 'Super Crunchers', p. 259
[7] http://www.google.com/googleblogs/pdfs/google_predicting_the_present.pdf

However these reports are only available with a lag; the data for a given month is generally released about halfway through the next month, and are typically revised several months later. Google Trends provides daily and weekly reports on the volume of queries related to various industries….We are not claiming that Google Trends data help predict the future. Rather we are claiming that Google Trends may help in predicting the present'. Their model is illustrated by examples from retail sales, automotive sales, the property market and the travel industry (predicting visits to a particular destination). While at this moment these methods stand at their infancy, and they need to undergo rigorous scientific scrutiny, my guess is that we'll see their explosive deployment in the very near future. The main attractiveness of these approaches is timeliness and (low) cost. In terms of timeliness they are at the cutting edge of what we can ever expect: *real time statistics.* This is the key feature of what I intuitively call Official Statistics 2.0, i.e. reducing the time gap between 'reference period' / 'occurrence period' and 'date of publication' to almost zero. I am aware that this may sound as a dream today but it will probably be the predominant--routine--characteristic of Statistics in the next decade.

At the time of writing this paper, I came across an overenthusiastic article[8] by Steve Lohr about the statistical profession, published in the New York Times (5 August 2009): *'For Today's Graduate, Just One Word: Statistics'.* The whole article is some kind of re-discovery of the importance of *statisticians* in a world dominated by information and communication technologies. I have selected a few interesting extracts from that article: *'The rising stature of statisticians, who can earn $125.000 at top companies in their first year after getting a doctorate, is a byproduct of the recent explosion of digital data. In field after field, computing and the Web are creating new realms of data to explore—sensor signals, surveillance tapes, social network chatter, public records and more. And the digital data surge only promises to accelerate, rising fivefold by 2012, according to a projection by IDC…'. '…We're rapidly entering a world where everything can be monitored and measured, said Erik Brynjolfsson, an economist and Director of the MIT's Center for Digital Business. 'But the big problem is going to be the ability of humans to use, analyze and make sense of the data'.*
The obvious potential of web-traffic data for statistical purposes is also highlighted in this New York Times article: *'It is the size of the data sets on the Web that opens new worlds of discovery. Traditionally, social sciences tracked people's behaviour by interviewing or surveying them. But the Web provides this amazing resource for observing how millions of people interact, said Jon Kleinberg, a computer scientist and social networking researcher at Cornell'.* There is as well a reference to Hal Varian, chief economist at Google, who underlined *'I keep saying that the sexy job in the next 10 years will be statisticians. And I am not kidding'.*

The above described emerging trends will certainly re-shape the market of statistics (and the statistical profession) and will inevitably impact on Official Statistics as well. I keep saying to my colleagues in Eurostat, the OECD and on every occasion I meet with NSIs that 'we', i.e. official statisticians, should be involved in this process now and not simply remain passive observers of the change. There are basically two axes on which official statistics should undertake (leading) initiatives: (i) legislation and (ii) development of a quality framework. It may sound somewhat strange to re-open the discussion on statistical legislation (in Europe) soon after the adoption of the (new) Statistical Law[9]! However, this shouldn't be a surprise to anyone since the recently adopted Regulation on European Statistics reflects and governs what I call Official Statistics 1.0, i.e. the traditional models of collecting, analysing and disseminating statistics, which-there is no doubt-will continue to dominate still for a long period. At the time of drafting the new Regulation it was not possible to imagine the scale and features of ICT developments that will bring a real revolution to statistics in the next decade. As we'll see in the following sections, the world is moving rapidly towards a digital life. Statistics in a digital world require new legal provisions which are not covered adequately by current statistical legislation. Think for example about issues governing data collection in a digital world and compare it with traditional survey data; in a digital world, data and metadata are produced (stored, transferred, compared, computed, etc.) at every moment an action is performed, with or without the agreement (or knowledge) of those concerned, often without the possibility for individuals to 'opt-out', without clear rules[10] about ownership, use and re-use of data, linkage of data, protection of privacy and so on. By contrast, in a traditional mode of data collection we (and the respondents) know very well (or, at least, fairly well) all the above issues together with our legal obligations.

[8] http://www.nytimes.com/2009/08/06/technology/06stats.html?_r=2&th&emc=th
[9] Regulation (EC) No 223/2009 of the European Parliament and of the Council on European Statistics, published in the Official Journal of the European Union L87/164 of 31.3.2009
[10] Even when such rules exist, their enforcement and monitoring tends to be poor

Similar drawbacks can be observed in terms of (absence of) quality standards for statistics derived in this emerging digital life. Looking for example at the analysis of internet traffic data, we know very little about the real content, coverage and quality of the data sources, their potential biases and their appropriateness to be used as sampling frames. Unless we (official statisticians) are deeply involved in the development of standards and quality frameworks for these new types of data and sources, there will always remain a risk that statistics produced by private actors will not necessarily meet the expectations of citizens for trustworthy and reliable information. Despite a few scattered efforts of Google to publicly disclose some algorithms and codes used for the calculation of their trends and indices, I can never be sure about their selectivity or not; to put it blindly, our knowledge today of what is behind the calculations of those who dominate the world of digital data is near to a 'black box'. This is a potentially serious situation--in terms of future developments in the statistics market--which necessitates appropriate actions to be undertaken by official statisticians along the lines described here. The risks of misuses of statistics in a digital world are similar or even greater than what has happened recently with the financial crisis.
Nassim Taleb, the well-known author of the international best seller 'The Black Swan', illustrates this aspect of misusing statistics in an original essay[11] entitled *'The Fourth Quadrant: A Map of the Limits of Statistics'.*
He underlines that '*Statistical and applied probabilistic knowledge is the core of knowledge; statistics is what tells you if something is true, false, or merely anecdotal; it is the "logic of science"; it is the instrument of risk-taking; it is the applied tools of epistemology; you can't be a modern intellectual and not think probabilistically—but.....Statistics can fool you. In fact it is fooling your government right now. It can even bankrupt the system (let's face it: use of probabilistic methods for the estimation of risks did just blow up the banking system)'.*
Taleb claims that *those who are putting society at risk are "no true statisticians", merely people using statistics either without understanding them, or in a self-serving manner.*

It is quite encouraging that at this moment there is a visionary leadership in the ESS who recognises the need to reshape the way we produce official statistics and reflect on a new professional paradigm.
In the February 2009 issue of *'Eurostat Infos'*[12]*,* the Director General of Eurostat, Walter Radermacher, has written an article about how he imagines the world of statistics in 2017 (*Rethinking the future*). He is emphasizing that 'new technologies have revolutionised the way we live, learn, work and play. New technologies have radically changed the ways in which people, industries, governments and society interact. And they are rapidly changing the environment in which *we develop, produce and communicate official statistics'.* He is addressing a central question which is of particular relevance to the market of official statistics: *' Will Eurostat and the ESS be able to fulfil our mission in ten years' time if we continue business as usual? Or will other data providers have taken over the market'*? The ideas and questions surrounding the vision of Eurostat's DG have been extensively discussed within the European Commission and the ESS, and have now led to the adoption of a Commission Communication[13] on *'the production method of EU statistics: a vision for the next decade'*. Implementing this vision will be a real challenge for the ESS. It is within this enabling framework that I expect official statistics to prepare themselves for their transition towards version 2.0.

What I'm trying to argue in this section is that in a 'digital society' where many of the actors in the statistics market will not be official statisticians, there will be an increasing need for a 'trusted third party' who will primarily be the guardian of principles, norms and standards of (official) statistics. It is the NSIs and supranational bodies like EUROSTAT, OECD and UNSD that can be entrusted this role. This need, i.e. for a 'trusted third party', will become more obvious in the next section where we look at future ICT developments presented in the form of possible scenarios.

2 Scenarios of a Ubiquitous world

It was a year and a half ago when I took over responsibility for 'statistics on the information society', and I must confess that the concept of 'ubiquitous computing' was unknown to me at that time. However, the concept is not new at all. It is nearly twenty years ago that Mark Weiser, a researcher at Xerox Palo Alto Research Center (PARC) wrote an article entitled *'The computer for the 21st Century'*,

[11] http://www.edge.org/3rd_culture/taleb08/taleb08_index.html
[12] Eurostat's newsletter available on its intranet website Cybernews
[13] COM(2009)404 final, Brussels, 10.08.2009. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0404:FIN:EN:PDF

published in *Scientific American*[14] in 1991, and subsequently in a series of other articles[15], where he illustrated his vision of a new stage in the history of human-computer interaction in which information processing will become part of the fabrics of everyday objects and activities. Weiser talked of ubiquitous computing as being a 'calm technology', a technology that resides everywhere in a way that you do not pay attention, in analogy to other great technologies in the history of mankind like 'writing and electricity'. '*The social impact of imbedded computers may be analogous to two other technologies that have become ubiquitous. The first is writing, which is found everywhere from clothes to labels to billboards. The second is electricity, which surges invisibly through the walls of every home, office, and car. Writing and electricity become so commonplace, so unremarkable, that we forget their huge impact on everyday life. So it will be with UC…Ubiquitous computing names the third wave in computing, just now beginning. First were mainframes, each shared by lots of people. Now we are in the personal computing era, person and machine staring uneasily at each other across the desktop. Next comes ubiquitous computing, or the age of calm technology, when technology recedes into the background of our lives'. (Mark Weiser)*

There is really a very extensive bibliography on all aspects and dimensions of ubiquitous computing. The aim of this section is not to define precisely all types of UC but rather to introduce a broader understanding of the concept, examine progress made so far, and look at possible future developments with a focus on the implications for official statistics and the ESS in particular.

A good summary of the issues involved in ubiquitous computing is very well presented by Anne Galloway in her paper *Intimations of Everyday Life: Ubiquitous computing and the city,* published in Cultural Studies[16] in 2004. '*In its broadest sense, ubiquitous computing is currently seen to comprise any number of mobile, wearable, distributed and context-aware computing applications. In this way, Ubicomp may consist of research into how information technology can be diffused into everyday objects and settings, and to see how this can lead to new ways of supporting and enhancing people's lives'.* Like many other authors and researchers in this field, Anne is putting particular emphasis on the implications of context-awareness and pervasiveness of ubicomp: *central to ubiquitous or pervasive technologies is the ability of computers to be perceptive, interpretive and reactive. In other words, information infrastructures must be able to shift from periphery to centre, and to recognize and respond to actual contexts of use. Context aware computing therefore relies on two types of information: physical location and user identity, both requiring extensive data acquisition, storage and delivery mechanisms.* It is actually these characteristics and key features of ubicomp that interest me as an official statistician: data, ownership of data, storage of data, meaningful analysis, disclosure of data, etc.

It is widely recognised today that the design of ubicomp systems is a matter of multidisciplinary collaborative research, involving engineers, urban planners, designers, sociologists, psychologists, lawyers, users, and many others who are concerned with aspects of 'every day life'. However, despite this recognition, the real involvement of scientists from humanities and social / cultural fields has been rather modest. As far as official statisticians are concerned, I am not aware of any active participation (not even a modest involvement) in the design of such systems. From the moment where ubicomp systems function on the basis of capturing, transmitting, linking, analysing (confidential) data, I really see an important place for official statisticians. Building a 'trusted information environment' should become a routine feature of the design of any ubicomp system; moreover, these technologies should also be designed in a way allowing the analysis of data for purely statistical purposes, replacing (and /or supplementing) several of today's data collections and therefore reducing response burden and costs, increase timeliness and capture emerging phenomena at an early stage. I will continue to argue throughout this paper for the need for official statistics to obtain their legitimate role in designing (data-rich) systems that will serve better the society of tomorrow.

Let's now see what has happened since the time the original ideas of Mark Weiser were presented. Has there been any progress? And what is likely to happen in the next decade? How certain can we be that UC and its promises will be realised at all?

While Mark Weiser would have expected a faster development and deployment of ubiquitous computing, there is no doubt that a lot of progress has been made, especially in terms of dedicated research projects and supportive political and regulatory environments which are necessary for a wider adoption of such systems. The first International Conference on Ubiquitous Computing was

[14] http://www.ubiq.com/hypertext/weiser/SciAmDraft3.html

[15] see for example: The coming age of calm technology, October 1996: http://www.ubiq.com/hypertext/weiser/acmfuture2endnote.htm

[16] http://www.purselipsquarejaw.org/papers/galloway_culturalstudies.pdf

organised in Karlsruhe, Germany in 1999. Since then, ubicomp conferences have become an annual event. Visiting the official website of the 11th conference, http://www.ubicomp.org , to be held in autumn this year, we can have a flavour of what is the state of art in ubicomp issues: *Ubicomp is the premier outlet for novel research contributions that advance the state of the art in the design, development, evaluation and understanding of ubiquitous computing systems. Ubicomp is an interdisciplinary field of research and development that utilizes and integrates pervasive, wireless, embedded, wearable and/or mobile technologies to bridge the gaps between the digital and physical worlds. Given its multi-disciplinary nature, Ubicomp has developed a broad base of audience over the past 10 years. Key audience communities are: Human Computer Interaction, Pervasive Computing, Distributed and Mobile Computing, Real World Modelling, Sensors and Devices, Middleware and Systems research, Programming models and tools, and Human Centric Validation and Experience Characterization.* Looking in the list of papers for this conference, we can observe that the most popular fields of implementation relate to health, mobility and energy / environment (this is also the case for other research projects). Just to mention a few of the titles: *Recognizing Daily Activities with RFID-based Sensors, Wearable Therapist: Sensing Garments for Supporting Children Improve Posture, Validated Caloric Expenditure Estimation using a Single Body-Worn Sensor, ViridiScope: Design and Implementation of a Fine Grained Power Monitoring System for Homes, etc.*

Like in so many other fields, developments and large-scale implementation of ubicomp technologies are 'applications-driven'. In particular, applications which improve the quality of our lives are more likely to facilitate the diffusion of these technologies at a faster rate. To illustrate the attractiveness of this, I would mention 2 additional recent examples. Infosys[17] has developed a managed information system named *Shopping Trip360* which provides shoppers with highly personalised and improved services, while manufacturers of CPG[18] products and retailers get unprecedented visibility regarding shoppers' behaviour and shelf activity. Amongst the several features of the system we find the following: (a) '*Smart Visual Merchandising* which is a rich user experience platform where shoppers can interact through their cell phones and get information on products while at the same time get recommendations on related products, and (b) perform '*Perpetual checkouts,* which allows shoppers to continuously ring items as they buy and pay automatically, thus avoiding lengthy checkout lines'.
The second example relates to a research project at the University of Southampton's Pervasive Systems Centre[19], which is called *Unobtrusive Welfare Monitoring System.* The aim of the project is to 'develop and demonstrate a pervasive welfare monitoring system using unobtrusive sensors for early detection and automated reporting of deteriorating physiological health parameters and enabling independent living'. Thinking for a moment of the long term demographic trends--with the segment of those aged 70 years and over representing the fastest growing part of the population--we can easily conclude that such systems will find several clients in the future. A critical issue in all debates of ubiquitous / pervasive technologies relates to ethics and privacy. Personal data on the health condition and health history of individuals in particular are highly confidential. Even if a system claims to be non-intrusive from a technical point of view, it may well be highly intrusive in terms of privacy. It is therefore possible to expect that users will require the adoption of adequate legislative safeguards prior to their acceptance of such systems. However, when we think of elderly persons with disabilities such safeguards might be of secondary importance (perceived as such by the individuals themselves); their primary concern is to protect their health and enjoy an independent living. These different perceptions of privacy may entail the risk that technology is massively deployed even without adequate privacy protection assurance, a risk that we should always bear in mind and try to avoid.

Another way to enhance our imagination and understanding of what might happen in the near future is to have a look at current EU-funded research. The EU is one of the most active and politically committed international actors in preparing Europe for tomorrow's digital society. There is a strong, over-the-horizon vision of the EU policies which is expressed not only financially but, as well as, through the preparation of a well balanced regulatory environment. A comprehensive, up-to-date and detailed description of the EU ICT Work Programme[20] for 2009-2010, can be found at the web site http://www.cordis.europa.eu/pub/fp7/ict/docs/ict-wp-2009-10_en.pdf. There are hundreds of collaborative and hyper-innovative research projects (ongoing or planned) to be funded by substantial

[17] http://www.infosys.com/ShoppingTrip360
[18] Consumer Packaged Products
[19] http://eprints.ecs.soton.ac.uk/16722/2/Unobtrusive_Welfare_Monitoring_System.pdf
[20] Updated Work Programme 2009 and Work Programme 2010, Information and Communication Technologies, *European Commission C(2009) 5893 of 29 July 2009.*

amounts of EU funds (more than 2 billion € of today's budget estimates). An important part of the projects (and the budget) is targeted towards designing our future internet; the research challenge is entitled *'Pervasive and Trustworthy Network and Service Infrastructures'*, and includes the following strands: 1. The Network of the Future, 2. Internet of Services, Software and Virtualisation, 3. Internet of Things, 4. Trustworthy ICT, 5. Networked Media and 3D Internet, and 6. Future Internet experimental facility and experimentally-driven research. I would strongly invite readers to navigate[21] through a number of these research projects in order to get a first-hand understanding of where technology is likely to be driving us in the next decade. For a comprehensive reading of all EU activities related to the future of the internet, it is best to visit the dedicated *foi portal* at http://www.future-internet.eu/ . Amongst the plethora of exiting and relevant information one can find on that website is the activities of the Future Internet Assembly (FIA) and the 2009 FIA book[22] entitled *'Towards the Future Internet – A European Research Perspective'*.

In May 2009, the European Commission published a report[23] entitled *'Future Internet 2020: Call for Action by a high level visionary Panel'*. It includes the visions of an industry expert group composed by 5 well known leading technology experts who describe the internet of 2020 in the form of scenarios. In their introductory remarks they highlight the tremendous differences between the current and future internet in saying *'…this future internet will be much faster and smarter, more secure, embracing not just information and content but also services and real world objects (things)….The internet was never designed for how it is now being used and is creaking the seams. We have connectivity today but it is not ubiquitous; we have bandwidth but it is not limitless; we have many devices but they don't all talk to each other. We can transfer data but the transfers are far from seamless……And, since security was an afterthought on the current Internet, we are exposed in various ways to spam, identity theft and fraud……By 2020 the Internet will be both laid out as public infrastructures and dynamically created by the objects connecting to one another. We need to see the Internet of the Future as this seamless fabric of classic networks and networked objects…'*.

The high-level panel of experts was invited by the European Commission to help prepare the ground for a new approach to PPP[24] (public private partnership) in this area. They describe their vision through illustrative scenarios and discuss (a) the reasons for which these scenarios are important for Europe, and (b) identify the *enablers* for each scenario. Their examples are given in a lively manner, and the reasoning behind these visionary ideas is derived from solid experience and rigorous scientific and research based evidence. I found very interesting (and highly probable to be realised soon) a scenario from the consumer sector, which assumes that the large scale adoption of RFID technologies and developments in the area of Internet of Things (or internet *with* things) will create new business models for a variety of personal services or *'Personal Mush-ups'.* Here is an extract from page 31 of the report:

> Anna loves shopping, especially since stores became personalised boutiques.
> As she steps into her local clothes shop Anna is recognised as a VIP customer – in other words, one who is prepared to share part of their personal data to get a loyalty card. As she gets closer to the racks of dresses tiny LEDs light up to indicate those dresses that would fit her size, with some colour coding to signal special discount, just for her. She goes to a dressing room where a mirror-like screen recognises her and presents her with a choice of dresses to try on, virtually. Knowing her size, the shop's computer can easily do the trick of dressing her image reflected by the screen. Anna finds one she likes and asks a sales assistant to get her the real thing.
> As she tries it out, the mirror reflects her image and starts proposing some accessories. Some of them are not sold at the store but are offered by other merchants. Her image becomes a mashup to advertise products. Anything that she is able to click through, by touching the mirror, will generate some sort of revenue for the shop.
> Anna's friends can join in too. With the new social retailing applications she asks her friend Katie to take a look and provide advice, even though she's not there in person. Her friend can access the mirror through her own devices, be it a television, or a mobile phone, or a navigator screen. Anna chats through the options, and since the dress has an Internet address Katie could even try it on (virtually) as well. Or she may wish to choose a different colour, a different fabric from the store inventory, or see how it fits and how it looks if they were to walk together to a party.

[21] http://cordis.europa.eu/fp7/ict/programme/challenge1_en.html
[22] Edited by G. Tselentis et al., IOS Press, 2009, ISBN 978-1-60750-007-0
[23] http://ec.europa.eu/information_society/activities/foi/library/epr.pdf
[24] A Communication from the Commission on *'A public-private partnership on the Future Internet'* is expected to be adopted by the EC in Autumn 2009.

There are various enabling mechanisms, events and initiatives at EU level that could be mentioned in this regard, and which demonstrate clearly that the world in 2020 will be largely ubiquitous, pervasive, and digital. If only a small fraction of the hundreds of research projects is to be translated into real life applications, even then, this would be sufficient for transforming our societies. Next to the support of the technical developments, the EU is also preparing the way for a proper regulatory framework, which is a highly 'delicate' and sensitive part of the whole technology package. Trying to balance the 'free internet' principle with the legitimate needs of citizens for 'protection of privacy and identity', protection of intellectual property rights, etc. is not an easy task. It is worth mentioning here a few EU actions that illustrate the political commitment of Europe in safeguarding the interests and shared values of its citizens while--at the same time--fostering cutting edge research into the future of ICT. For example, in order to support the definition of the previously mentioned PPP on the future internet, the Commission has just launched a large scale study[25] entitled *'The economic and societal impact future internet technologies, services and applications will enable Europe and elsewhere- A quantitative study 2015-2020'*. The focus areas will be in smart energy grids, smart environmental information systems, transportation and mobility, and smart healthcare systems. Another example relates to the adoption of a Communication on the 'Internet of Things – An action plan for Europe[26]'. While the Communication highlights the expected economic and societal benefits, it is more focused on issues of governance, standards, identification of emerging risks, protection of personal data, and the citizens' 'right to silence of chips'. The European Commission has also issued a specific Recommendation[27] on the implementation of privacy and data protection principles in applications supported by RFID on 12 May 2009. These initiatives are part of a wider and consistent ICT policy which had already been initiated a few years ago within the i2010[28] framework, which is currently being renewed (extended) for the period 2010-2015. DG INFSO is organising a *public hearing* on the *post-i2010 strategy* which will be held in Brussels on 23 September: http://ec.europa.eu/information_society/eeurope/i2010/pc_post-i2010_hearing/index_en.htm .

As final piece of supporting evidence for my conviction that we are transiting towards a ubiquitous and pervasively interconnected world I would like to site the book of Adam Greenfield *'Everyware[29] – The dawning age of ubiquitous computing'.* In a series of short essays (thesis), Adam explains in a fascinating manner how *everyware is already reshaping our lives, transforming our understanding of the cities we live in, the communities we belong to and the way we see ourselves.* To the question 'When do we need to begin preparing for everyware', Adam's (well argued) response is 'Now': *…However long it may take a full-fledged everyware to appear, the moment to begin developing a praxis appropriate to it is now*[30]*.* Official statistics cannot escape from the impact of *ubiquitous computing or 'everyware'.* We have seen that--by definition--the concept of ubiquitous computing consists of data, metadata, and information processing embedded in the objects, body, surfaces and (through them) in the activities of everyday life. While there are several technical and methodological challenges for official statistics to tackle in this emerging world of ambient informatics and context aware technologies, there are as well some great responsibilities not to be underestimated: developing quality frameworks with internationally agreed standards for *real-time* statistics, guiding consumers and prosumers of statistics to avoid confusion in an environment of *'too much information'*, helping legislators to adopt appropriate, operational and comprehensive rules and measures for the governance of highly confidential and relational[31] data environments and assist engineers and private data-source managers in the design of 'trusted information systems' are just some examples of areas of competence and responsibility of official statistics.
I hope that by now we can agree with three basic observations: (i) while it might be difficult to predict the exact timing and scale of ubiquitous computing there is adequate evidence suggesting that it will rather happen sooner than later; in particular, political willingness, industry's involvement and the scale of research activities advance any discussion of ubiquitous technologies and their deployment to a

[25] http://ted.europa.eu/Exec?DataFlow=N_one_doc_access.dfl&Template=TED/N_one_result_detail_curr.htm&docnumber=239608-2009&docId=239608-2009&StatLang=EN
[26] COM(2009) 278 final, Brussels, 18.06.2009.
[27] http://ec.europa.eu/information_society/policy/rfid/documents/recommendationonrfid2009.pdf
[28] http://ec.europa.eu/information_society/eeurope/i2010/index_en.htm
[29] Published by New Riders, 2006, ISBN 978-0-321-38401-0
[30] A. Greenfield, Everyware, p. 181
[31] Limitless possibilities for matching and 'relating' data and information, empowered by dynamic models. Such environments, while they represent obvious benefits for statistical analyses, they also entail high risks of breaching confidentiality and privacy.

new decisive phase; (ii) ubiquitous computing will reshape many things in our life. It will impact on almost everything. Official statistics will not be exempt from this. There is a responsibility and legitimacy for official statistics to be actively involved in the design of 'trusted information systems'. OS will become a 'trusted third party 'in this context; (iii) the moment to start developing OS 2.0 is now.

3 Some domains to explore further with pilot (statistical) projects

There are several areas in which official statistics could embark with research projects and pilot work to investigate the potential of using pervasive technologies for statistical purposes. For illustration purposes I will briefly describe three such areas. The selection of the areas is driven by a number of considerations (e.g. technical advancements, policy commitment, investment by industry, etc) which, in my view, provide adequate justification for their inclusion here. My hope is that we can take some of these ideas further and transform them into concrete research / pilot projects during the next couple of years.

3.1 Using the Internet as a Data Source

The rationale behind this selection is manifold. First, internet is becoming an extremely rich data source that can reduce substantially the response burden on individuals and businesses[32]. Second, internet is the predominant network infrastructure of tomorrow's digital world; we have seen in the previous section that internet is currently attracting large funding for cutting edge multi-disciplinary, multi-stakeholders / international research projects. EU policy and legislation are seriously making progress by creating an 'enabling' and stimulating environment for the development of the 'future internet'. Third, we already have some early experiences and pilot initiatives in exploring the use of the internet as a data source for statistical purposes. We can actually build on these experiences and launch more extensive testing, experimentation and evaluation of various alternatives.

It is worth mentioning here the original work carried out by *Dialogic* on behalf of the Netherlands Ministry of Economic Affairs[33] through a research project entitled *'Internet as a Data Source (IAD)'.* The outcome of the project is described in the publication *'Go with the dataflow'* and it is downloadable from the website of the ministry http://www.ez.nl/dsresource?objectid=157266&type=PDF . The research questions of the project were (a) *to identify new data and indicators derived directly from the Internet and describe new phenomena associated with EDE (emerging digital economy), and* (b) *to explore and assess the usefulness of the various IaD methods for deriving new, extra and substitute data for the EDE.* These questions are addressed through 8 case studies covering both the 'old economy' (established markets & phenomena / low levels of digitalization) and the 'new economy' (emerging markets & phenomena / high levels of digitalization). Various IaD methods are tested and evaluated in terms of their statistical *pros and cons,* as well as in terms of their *usability and disadvantages*. The typology of the experimented methods included three strands of measurements: *(i) User-centric measurements* (spyware, traffic monitoring at operating system level), *(ii) Network-centric measurement* (Deep packet inspection), and *(iii) Site-centric measurement* (spiders). As it has been stated earlier, the current state of these methods (in terms of statistical robustness) is very primitive and we need to tackle several issues of concern: methodological, regulatory (legal) and privacy being the most important. However, these concerns, while taken seriously, should not prevent official statistics from taking a forward looking approach through additional collaborative research and experimentation in this domain. The Dutch IaD study includes a section on policy recommendations in which there is a 'plea' addressed to NSIs to play the function of a *clearing house* for Internet-based statistics. In its final words, the report invites publicly funded statistical agencies as well as Eurostat and the OECD to develop a roadmap for innovative methods and innovative statistics.

The ideas of a *clearing house* function or of a *trusted third party* (which is my preference) or the *'quality labelling' / 'certification'* approach (proposed by the OECD), while being different from an operational viewpoint, they all point to the same need: *developing a quality framework for Internet-based data*. I am confident that in the framework of the recently established expert group on *'Measuring network activity & behaviour + Synergies between official & private ICT data',* which will report to the OECD Working Party on Indicators for the Information Society (WPIIS), we will identify

[32] During the last 5 years the European Statistical System has been under severe pressure to reduce the response burden (particularly for enterprises) due to statistical inquiries

[33] The main report, its annexes and information about the consultation process can be found on the website of the ministry: http://www.ez.nl

the terms of reference for a number of research projects which would involve official statisticians, academics and experts managing private ICT data sources.

3.2 Automated data collection for price statistics

The rationale behind the selection of price statistics (price indexes) is driven by discussions and advances in RFID technology and its deployment in the consumer products sector in particular. The huge market potential from the expected near-future growing deployment of RFID-based technologies is highlighted in a (February 2009) Deutsche Bank Research[34] paper entitled '*RFID chips – Enabling the efficient exchange of information'.* The paper underlines that *'….while turnover is likely to increase by an average of 19% p.a. in Germany between 2006 and 2016, the pace may reach as high as 25% p.a. worldwide'.* We have good reasons to believe that the roadmap for addressing the outstanding political and technological challenges is well on its way of implementation. Worries about costs of RFID tags in particular seem to diminish since the predictions are now set to as low as 0,01 € per tag. Moreover, research into the so-called *'printable electronics'* is making printable RFID a reality. Basically, printable RFID, also referred to as "organic" or "plastic" RFID, is a method of producing tags by printing electronic inks that have characteristics similar to those found in the standard silicon microchips. This method is expected to reduce costs dramatically while it increases manufacturing speed and precision.

It is therefore reasonable to expect that the barcode technology will (soon) be replaced or largely substituted by RFID-based technologies. Let's now see what this may imply for price indexes. One of the basic advantages in using RFID-collected prices is that data are based on actual transactions rather than price quotations and therefore such measurements reflect more accurately variations in what is being measured. Moreover these data can provide almost a census of all transactions (elimination of sampling error or drastic reduction due to the use of much larger data sets) rather than a sample and can be collected continuously rather than at infrequent fixed intervals as it is the case with traditional retail price collection. RFID tags can host much more detailed data than 'price and type' of product, which are essential attributes of the quality of the item being measured. Such information can be used to make *quality adjustments* that are necessary to assure precise measurement of changes in the cost of living. Another advantage from RFID-based systems is that the measurement of 'price and quantity' takes place simultaneously (while conventional surveys for prices and revenues are carried out separately and at different frequencies, resulting in substantial lags with regard to data availability). Overall, the concurrent observation (and collection) of prices and revenues will potentially imply significant reduction of costs and burdens for both the NSIs and the responding units; it will also yield more consistent measurements of price and revenues and it will improve data quality while the usefulness and richness of these data bases for research will be greatly enhanced. The fact that this technology allows for continuous price (and quantity, quality) observations, it will make possible to capture the average prices paid by consumers more accurately (eliminating fixed-time data collection effects; e.g. prices sampled at fixed times might not be representative of what consumers pay on average, since a lot of price variations can be observed during a day, week or month, during promotions, sales, etc). Last but not least, RFIDs can cope with early identification of new products and new shops and emerging changes in consumer patterns (it is well known that the slow adjustments for new products and shopping patterns are a source of mismeaurement of inflation by CPIs).

While I have been talking only of possible advantages of using RFID-based price data, I am aware that there are formidable challenges and technical, methodological and practical issues to consider as well. These challenges and concerns are similar to those that have been identified by a number of research projects dealing with the use of scanner-data for the construction or improvement of CPIs. NSIs can build on these experiences and launch similar research and experimentation in the potential of using RFID-based price collections. I am confident that the results and knowledge to be acquired from such research will be beneficial not only for price indexes' statistics as such but for a wider range of statistical measurement and analyses. To mention just one of the statistical challenges of ubiquitous computing which is typical for RFID environments and which will require the NSI's active involvement: how to cope with the explosion of data and information? New models are needed for enabling official statistics to translate continuously and dynamically the petabytes of data into useful and meaningful statistics. Experimentation with observing and analysing millions of price and transaction data for numerous products sold in hundreds or thousands of shopping establishments located in many geographical regions will certainly provide us with valuable expertise in this regard.

[34] http://www.epcglobalinc.org/home/Deutsche_Bank_RFID.pdf

3.3 Digital healthcare may re-shape health statistics

Healthcare is probably one of the major sectors in which ICT are massively and pervasively present both in terms of current deployment but also as far as experimentation for future developments is concerned. Of particular interest to official statistics is that deployment of pervasive and context aware technologies for health and healthcare purposes, implies that an ever increasing amount of health-related individual data are produced, stored, shared, transferred, consulted, analysed, linked with other information, etc. Questions of ethics, confidentiality, privacy and data protection in general are of paramount importance for the good functioning of such systems. The members of the ESS have long experience and a good record in dealing with confidential data. Throughout the years the ESS has proven its ability to act as a trusted and credible body in collecting, maintaining and analysing confidential data (including health related data) for statistical purposes.

I would argue that official statistical agencies have the legitimacy to belong to those groups of actors for whom access to (anonymised) electronic health information is permitted for purposes of statistical analyses. Several countries are introducing systems of electronic health records (EHR) where the health history of individuals is stored and updated on every occasion of a medical consultation, medication, treatment or other health-care related action is taking place. While it will take many years for a full fledged EHR system to be implemented in all Europe, once established it will constitute a *paradise for health statisticians.* Longitudinal health records, matched with other socioeconomic characteristics will eventually provide tremendous possibilities for research and analyses. I can imagine that developing such systems in an internationally comparable context is not an easy task. Agreement on the adoption of international standards for recording health information, development of common interfaces, agreeing on the set of common data, defining quality standards (labelling / certification) of data coming from third countries and / or non-official sources, etc are some of the issues that should be considered.

What types of pilot projects can be launched with the involvement of official statisticians? I would suggest that countries which have already started working with the digitalisation of the healthcare sector (e.g. Denmark) can take the lead in investigating the potential future uses of such systems for statistical purposes. In particular we can examine the role of EHR in substituting current (health interview) surveys which include ‘short’ or ‘long’ modules on the (perceived) health status of the population. Comparisons of data derived from EHR with those coming from health examination surveys can also be included in the research objectives of such pilot projects.

The title of this section could have been written the other way round: *'Statistics are reshaping health and healthcare'.* This function of official statistics has always been performed with outmost professionalism and credibility; while NSIs will continue to fulfil their mission in providing a sound basis for policy decisions in the health sector, it is legitimate to say that time has come for the health sector--in its digital and ubiquitous new look--to enable official statistics benefit from the emerging large scale digitalisation. In return, official statistics will provide an ever increasing high quality information service.

4 Concluding remarks

This is an awareness raising paper. A profound technology shift is happening now that requires changes to our professional paradigm and our business model. Ubiquitous computing and context aware technologies will reshape many things in our life and it will impact on almost everything. Official statistics will not be exempt from this.

The market of official statistics is becoming increasingly attractive to private actors. 'Google queries' and other types of internet traffic flows represent new challenges for official statistics. Many of these developments are taking place outside the usual environment of official statistics and they raise obvious questions regarding quality and relevance. In a future digital society where many of the actors in the statistics market will not be official statisticians, there will be an increasing need for a 'trusted third party' who will primarily be the guardian of principles, norms and standards of (official) statistics. It is the NSIs and supranational bodies like EUROSTAT, OECD and UNSD that can be entrusted this role.

While it might be difficult to predict the exact timing and scale of ubiquitous computing, there is adequate evidence suggesting that it will rather happen sooner than later. A ubiquitous and pervasive world will require official statistics to change towards their version 2.0; the moment to start preparing for this is now. This is both a challenge and a responsibility.

Technological advances, political commitment, industrial investments, multidisciplinary research and visionary leadership in the ESS create a propitious momentum for official statistics to get involved in a number of pilot projects that will pave the way for a timely transition from OS 1.0 to OS 2.0. A major challenge and responsibility for official statistics will have to do with the management of petabytes of data and information explosion. The father of ubiquitous computing, Mark Weiser, promised that ubiquitous computers will handle this issue. Official statisticians have to prove that this is feasible.

> *'Most important, ubiquitous computers will help to overcome the information overload. There is more information available at our fingertips during a walk in the woods than in any computer system, yet people find a walk among trees relaxing and computers frustrating. Machines that fit the human environment, instead of forcing humans to enter theirs, will make using a computer as refreshing as taking a walk in the woods'* [Mark Weiser]

5 Acknowledgments

I would like to thank my colleagues in Eurostat Costas Giannakouris, Rik Storm and David Thorogood for spending their precious time in reading this paper and providing thoughtful comments.

6 References

1. Ayres, I.: *Super Crunchers,* Bantam Dell, New York, 2007, ISBN 978-0-553-38473-4
2. Choi, H. and Varian, H.: *Predicting the Present with Google Trends,* 10 April 2009, http://www.google.com/googleblogs/pdfs/google_predicting_the_present.pdf, last accessed on 10 August 2009.
3. Commission of the European Communities: *Recommendation on the implementation of privacy and data protection principles in applications supported by radio-frequency identification,* Brussels, 12.5.2009, C(2009) 3200 final.
4. Commission of the European Communities: *Communication from the Commission, Internet of Things – An action Plan for Europe*, Brussels, 18.6.2009, COM(2009) 278 final.
5. Commission of the European Communities: *Updated Work Programme 2009 and Work Programme 2010, Information and Communication Technologies,* C(2009) 5893, 29.7.2009.
6. Commission of the European Communities: *Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade,* Brussels, 10.8.2009, COM(2009) 404 final.
7. Commission of the European Communities: *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A public-private partnership on Future Internet,* planned for adoption in autumn 2009.
8. Deutsche Bank Research: *RFID chips – Enabling the efficient exchange of information,* February 2009, http://www.epcglobalinc.org/home/Deutsche_Bank_RFID.pdf, last accessed on 14 August 2009.
9. Dialogic: *Go with the dataflow! Analysing the Internet as a data source,* The Hague, April 2008, available at http://www.ez.nl, last accessed on 28 August 2009.
10. European Commission: *Future Internet 2020 – Call for action by a high level visionary panel,* Brussels, May 2009, ISBN-13-978-92-79-11320-8.
11. Galloway, A.: *Intimations of Everyday Life: Ubiquitous computing and the city,* Cultural Studies, 2004, http://www.purselipsquarejaw.org/papers/galloway_culturalstudies.pdf, last accessed on 10 August 2009.
12. Greenfield, A.: *Everyware – The dawning age of ubiquitous computing,* New Riders, 2006, USA, ISBN-13: 978-0-321-38401-0.
13. Infosys: *Shopping Trip360,* http://www.infosys.com/ShoppingTrip360, last accessed on 28 August 2009.
14. Lohr, S.: *For Today's Graduate, Just One Word: Statistics,* New York Times, 5 August 2009.
15. OECD: *Building synergies between official statistics and private source data,* DSTI/ICCP/IIS/RD(2009)2, Paris, 21.4.2009.
16. Official Journal of the European Union: *Regulation (EC) No 223/2009 of the European Parliament and of the Council on European Statistics,* L87/164 of 31.3.2009

17. Taleb, N.: *The Fourth Quadrant: A Map of the Limits of Statistics,* published at the Edge, http://www.edge.org/3rd_culture/taleb08/taleb08_index.html, last accessed on 10 Aug. 2009.
18. Tselentis, G. et al editors: *Towards the Future Internet – A European Research Perspective,* IOS Press, 2009, ISBN 978-1-60750-007-0.
19. Weiser, M.: *The computer for the 21st Century,* Scientific American, 1991, available at the site: http://www.ubiq.com/hypertext/weiser/SciAmDraft3.html last accessed on 10 August 2009.
20. Weiser, M.: *The coming age of calm technology,* October 1996, available at the site http://www.ubiq.com/hypertext/weiser/acmfuture2endnote.htm last accessed on 10 August 2009.