Global Conference on a Transformative Agenda for Official Statistics Towards a Strategic Framework for Statistics in Support of the Post-2015 Development Agenda

Co-organized by the United Nations Statistics Division (UNSD) and the Statistical Office of the European Communities (Eurostat)

United Nations Headquarters, New York,
Conference room 5
15-16 January 2015

Session 2: Innovations in data collection, data dissemination, data access and data analytics

Modernisation: Evolution or revolution

Prepared by Pádraig Dalton, John Dunne and Donal Kelly

Global Conference on a Transformative Agenda

15-16 January 2015 New York

Session 2: Innovations in data collection, data dissemination,

data access and data analytics

Modernisation: Evolution or revolution

Prepared by Pádraig Dalton, John Dunne and Donal Kelly, Central Statistics Office, Ireland

I.   Introduction

The recent report of the UN secretary-General's Independent Expert Advisory Group on the Data Revolution for Sustainable Development entitled "A World That Counts: Mobilising the Data Revolution for Sustainable Development Goals"[1] provides a very clear and concise articulation of the value of Official Statistics but also the challenge facing Official Statistics:

> *Data are the lifeblood of decision-making and the raw material for accountability. Without high-quality data providing the right information on the right things at the right time; designing, monitoring and evaluating effective policies becomes almost impossible.*

The world in which we live is changing fast and indeed is changing continuously. Operating in such an environment, as Official Statisticians do, requires an ability to adapt and innovate if we are to remain relevant. The pace of change in recent times has created many challenges for Official Statisticians but also a range of opportunities. Our focus must be on the "art of the possible" while remaining cognisant of, but not paralysed by, the challenges.

Modernisation can take many forms and in the current environment some of the issues that immediately spring to mind include Big Data, developments in technology, role of social media, impact of open data, the potential of artificial intelligence, standards based modernisation and data visualisation. However a broader view of modernisation is required given the variety of starting points from which compilers of Official Statistics are setting off.

Some of the relevant issues include the fact that there are those operating in an environment where the fundamentals of an independent, objective and impartial statistical system are not enshrined in law or practice, and/or the right of access to administrative data is not guaranteed, the data infrastructure required to exploit the vast quantity of available secondary data is not in place, the skill-sets required to engage with these new data sources are not available etc.

---

[1] http://www.undatarevolution.org/report/

If we are to provide trusted, reliable, independent and internationally comparable data to support initiatives such as Sustainable Development Goals then many of these issues will need to be addressed.

The pace of change needs careful consideration and indeed looking at some of the drivers for change it is clear that in some specific areas there is a particular urgency attached to the modernisation agenda. It is also clear that there are other areas where change can only be affected over a longer time horizon. So one of the questions to be looked at is in what areas do we need to revolutionise and where do we need take an evolutionary approach?

The focus of this paper is not to list all the modernisation initiatives or technological developments that may impact on our world but to highlight the high-level strategic challenges facing the official statistical community in both developed and developing countries.

The paper starts with a brief synopsis of the environmental context within which Official Statisticians operate. We then go on to look at the need for a Modernisation Maturity Model (MMM) which could provide a framework within which to assess the maturity of individual statistical organisations, but also the maturity of the global statistical community, from a modernisation perspective. Some specifics are then examined in relation to data sources, institutional and legal frameworks, technological developments, the role of standards in modernisation programmes, communications questions and finally a brief conclusion.

II.   Environmental Context

Some of the significant environmental issues influencing our thinking at the moment include the increasing global nature of official statistics and the focus on globalisation as a statistical challenge, the Digital Age and the emergence of new technology and new data sources (Data Deluge), the privacy / data security debate, the increased level of scrutiny of official statistics, the politicisation of official statistics, and value for money issues.

Increased User Demand: The user community of official statistics has changed. The economic crisis has heightened awareness of our products and we are now faced with a broader, more demanding and informed cohort of users. This is undoubtedly a positive development but it also brings a range of challenges that need to be addressed. Not only is there an appetite for a broader range of statistical products but users are also looking for a new range of dissemination channels.

Global Users: We have also seen the emergence of the "global user" where the focus is on international comparisons. The globalised world now more than ever demands data that are internationally comparable. Standards across a range of areas are of course central to the achievement of this objective and undoubtedly work remains to be done in this area.

The Digital Age: The Digital Age has brought new technology, new data sources and new challenges. Technology is constantly evolving and keeping pace with the level of change is difficult. New technology requires new skill sets and many National Statistical Institutes are

debating whether or not to grow the expertise within or to outsource. This is an important debate especially in the context of growing statistical demands and a greater focus on value for money.

Where should we be focussing our resources and are there opportunities to develop strategic alliances (including private sector organisations) in key areas e.g. IT?

Skills: The issue of acquiring the appropriate skill sets across a broad range of areas is particularly important. In a recent blog from Gartner[2] relating to Hadoop[3] deployment the shortage of Hadoop skills was referenced and the challenges of creating a working Hadoop environment even in advance of processing data in such an environment was highlighted.

If as Gartner suggests the skills required to even establish a Hadoop production environment are scarce, even for financial services companies with deep pockets, how can we as public sector organisations address this challenge? Perhaps this is an area where we should consider pooling resources?

Politicisation of Official Statistics: The increased politicisation of official statistics was perhaps inevitable following the recent global economic crisis. The relevance and role of official statistics became apparent to all and with official statistics now being increasingly used to measure performance the pressure placed on some compilers of official statistics has increased. Public trust is essential for any statistical system and the key building blocks are independence, objectivity in analysis and a commitment to maintain the integrity and confidentiality of data provided. This is a significant modernisation challenge in some areas and if we are to produce trusted, robust and internationally comparable data we must not lose sight of this challenge.

Resources: Value for money has always been a focus of official statistics and in recent years the drive to do more with the same, or in most cases, less has yielded efficiencies across the system. However, there is an inherent contradiction in extolling the virtue and need for high quality, timely and relevant data by policy makers and Governments while at the same time reducing the level of resources made available to the compilers of official statistics. Undoubtedly value for money must be a core objective for any statistical organisation but the increased demands and the new environments in which official statisticians now operate requires investment if we as a community are to meet the needs of our broad range of users.

When it comes to resources, as a community, are we too passive in how we deal with the unabated growth in the demands for information and do we have a consistent message to deliver as to how we should be addressing these demands as a global community?

---

2 http://blogs.gartner.com/nick-heudecker/hadoops-achilles-heel-in-2015/

3 Hadoop is often referenced as the technology of choice in relation to Big Data

III.   Institutional, legal and ethical

The key building blocks of any statistical system are independence, objectivity in analysis and a commitment to maintain the integrity and confidentiality of data provided. The achievement of these objectives requires an appropriate institutional setting embedded within an appropriate legal framework. For many countries the absence of a stable functioning statistical framework is the biggest modernisation challenge being faced.

This is particularly important in the current environment where as a global community we are moving towards an ambitious project to develop a new suite of Sustainable Development Goals for which official statisticians will be required to provide trusted, reliable, independent and internationally comparable data.

Pillars of an effective statistical framework

There are those that believe that there are three key pillars required to support an effective statistical system which provides the necessary independent and objective data that, as outlined in the "World That Counts" report, "are the lifeblood of decision making and the raw material for accountability".

- Data Protection which is designed to safeguards an individual's right to privacy (not an absolute right) and is applicable to the use of private information held by public sector organisations;

- Freedom Of Information which is designed to support open and transparent government through obliging public bodies to publish information on their activities and facilitating citizens to view and access any information held on them by those bodies; and

- Statistical legislation which enshrines in law the political and professional independence of NSI's along with the powers to collect information to compile statistics to inform society;.

The realisation of this framework even in the short to medium term for some is extremely ambitious. The focus at a minimum should be on the creation of the appropriate statistical framework in all developed and developing countries but can we develop a robust statistical framework in the absence of Freedom of Information and Data Protection frameworks?

Privacy, security and confidentiality

With the availability of huge volumes of data from new public and possibly private administrative data sources, and the need to create new statistical indicators from linked data sources, there will be many challenges in the areas of data security, privacy and confidentiality.

In the next section the issue of access to administrative data, amongst other things, is discussed. Indeed there will also be difficulties in accessing some Big Data sources. A

common theme relevant to the discussion on data sources is the issue of privacy. The so-called 'data deluge' brings with it many complex questions about privacy and data protection. Kitchin (2014) suggests that the 'privacy landscape is in flux', with some commentators suggesting that the battle for privacy has already been lost but others arguing that privacy is an 'indispensable structural feature of liberal democratic political systems'[4]. Kitchin argues that there is a broad consensus that existing privacy laws are no longer fit for purpose and need to be revised in line with the realities of modern living.

The challenge is not just a legal one even where the intention is purely to use the information for statistical purposes. Data sharing and linking creates the spectre of 'big brother' in the public mind and official statisticians must guard against anything that would damage this reputation or impact its ability to produce trusted statistics.

Privacy-efficiency trade-off

There is of course another angle to the privacy debate that needs to be considered in the context of greater access to administrative and Big Data sources. MacFeely and Dunne in their 2014 paper entitled "Joining Up Public Service Information: The rationale for a national data infrastructure"[5] raise an important issue relating to privacy, namely the "privacy-efficiency trade-off". Essentially they question whether there is an inherent contradiction between the demands for more efficient and cost effective public services, and increasing concerns over the privacy and protection of individual information. They note that:

> "Both at home and abroad, concerns over privacy and protection of individual information are live and increasing. Concerns such as the activities of UK Government Communications Headquarters, the US National Security Agency and 'information-rich' multinational enterprises have all prompted reaction, ranging from the establishment of the UK civil liberties group Big Brother Watch to the Data Protection Commissioner in Ireland conducting a 'privacy' audit of Facebook Ireland Ltd (Office of the Data Protection Commissioner, 2011). In Ireland, where privacy is highly valued, this is a particularly culturally sensitive issue (OECD, 2008, p. 211). Yet at the same time there is an appetite for increased public sector efficiency and a growing intolerance towards the administrative burdens imposed by government departments and state agencies on individuals and enterprises. The necessary trade-off between these two positions, however, does not appear to be well understood".

So what is the challenge facing official statistics in relation to privacy? Is it more than a legislative challenge i.e. a cultural challenge? Privacy engineering is an emerging field that NSIs will have to sponsor but also be seen to champion in order to maintain/enhance trust in Official Statistics? Either way it would appear at a minimum we have a significant communications challenge on our hands.

4Cohen, Julie. E - http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2175406

5 http://www.ipa.ie/pdf/Forum_Vol_61_4.pdf

IV.   Data Sources

For most statistical organisations data sources fall into one of three categories:

- Survey;
- Administrative including some Big data sources; and
- Other secondary data sources including some Big Data sources.

The hot topic from a data sources perspective at the moment is Big Data but in chasing and focussing on this particular issue, "Big Data", are we running the risk of missing something more important?

Big Data is very much in an evolutionary phase. As a community we are still trying to identify the specific Big Data sources that can and will support the production and dissemination of Official Statistics. The use of administrative data for statistical purposes is well established and has been "main-streamed" in the systems of many NSIs.

Access to administrative data for NSIs is not universal and is an issue that needs to be addressed. While the global statistical community is discussing the potential of Big Data for the production of official statistics the use of administrative data for the actual production of official statistics is well established.

Of course the argument here is not that we should ignore Big Data and focus exclusively on administrative data but rather reflect on the fact that the issue of access to administrative data, where the benefits for the production of official statistics are known, has not yet been resolved.

There are those however that are well positioned to lead the way on Big Data. The potential of 'Big Data[6]' to support the development of new statistical outputs has received considerable attention. The opportunity to harvest existing data sources while reducing dependence on traditional and more expensive sources, reducing costs, improving coverage and possibly timeliness has been highlighted by many.

Moving from "potential" to actual delivery of official statistics must be the next step. There are a range of initiatives underway examining the factors associated with the use of Big Data sources for official statistics at national and international level in areas including smart metering data, web scraping and mobile phone data.

While the benefits of Big Data have been highlighted there are those that caution against "conflating data volume with insight, utility and value"[7] and argue that the expertise and knowledge built up over many years using tried and trusted methods cannot be discarded or

---

6 "Big data is a buzzword, or catch-phrase, used to describe a massive volume of both structured and unstructured data that is so large that it's difficult to process using traditional database and software techniques. In most enterprise scenarios the data is too big or it moves too fast or it exceeds current processing capacity"(http://www.webopedia.com/TERM/B/big_data.html).

7 Kitchin, R. and Lauriault, T. (2014) The Programmable City Working Paper I. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376148 [September 2014]

replaced overnight. Similarly, others have argued that a common fallacy of Big Data is an expectation that more data results in more information and better insights.

The reality, as always, is perhaps somewhere in-between. Big Data is not necessarily better than traditional data sources and the value of the insights to be gained depends on the nature of the data itself, the ongoing availability of the data, availability of the IT tools for analysis, availability of the specific skill sets necessary to mine it (as discussed earlier in the paper) and the ability to address the resource implications in managing Big Data.

As a global community are we actively managing the expectations surrounding Big Data in advance of getting greater clarity on its actual usefulness for official statistical purposes? The "World That Counts" report tends to avoid the term Big Data and instead focus on "new and traditional data sources" which is perhaps a useful way of avoiding the excesses of the hype surrounding Big Data.

V.   Technological change

There have been major technological advancements in recent years that have the potential to significantly alter how we do our business from the proliferation of mobile technology to improvements in data visualisation. One of the most significant challenges is to try to maintain pace with developments in the digital world.

Over the past five years the widespread proliferation of internet-enabled mobile devices[8], better access to broadband and the explosion of social media use, have changed the way people interact with one another and changed dramatically the volume of data that is continuously generated about the lives of individuals. Similarly, developments in smart technologies mean that practically anything can now be digitally-enabled and monitored resulting in a stream of data from the 'internet of things'[9]. In the business world, companies of all sizes are becoming more data aware and are investing heavily in collecting and analysing huge volumes of data to help them make better business decisions.

This challenge has been recognised by the ESS, and the European Statistical System Vision 2020 document points to the need for further collaboration and the sharing of expertise between national statistical institutes and the need for public-private partnerships in order to develop more efficient production methodologies and in doing so creating capacity.

A key question is how do we come together to exploit the developments in technology and how do we learn from each other's experiences? How can we engage with third parties (both public and private) without undermining our, real or perceived, privacy reputation?

Considering the pace of technological advancements, and of course the ongoing challenge of producing official statistics, is it unrealistic for individual statistical institutes to be at the

---

8 An industry survey estimates that 86% of Irish adults now have access to a device that gives them mobile internet access (eircom Household Sentiment Survey, March 2014)

9 "The Internet of Things (IoT) is the interconnection of uniquely identifiable embedded computing devices within the existing Internet infrastructure". http://en.wikipedia.org/wiki/Internet_of_Things

cutting edge of technological development? Should NSIs be software and applications developers or should we focus on the development of specifications to agreed international standards? If we are to collaborate in a public-private partnership type framework should we be doing this at national level or as a statistical community do we first need to come together and agree a cross-national approach/position?

VI. Standards based modernisation

The importance of standards based modernisation is evident from all our experiences in official statistics. At international level there are a multitude of actors working in this particular domain to develop, agree and implement standards.

The work being undertaken by the UNECE High Level Group for the Modernisation of Statistical Production and Services (HLG MOS)[10] is particularly relevant. The HLG believes that countries can achieve greater efficiency in the production of many statistical products by using common language and standards, such as the Generic Statistical Business Process Model (GSBPM), the Generic Statistical Information Model (GSIM), Data Documentation Initiative (DDI) standards, Statistical Data and Metadata Exchange (SDMX) and the Common Statistical Production Architecture (CSPA) to describe national statistical systems. It is hoped that international synergies and economies of scale can be achieved when similar processes and structures exist in every country, as new innovations or methodologies developed by one country can be easily applied to another.

The globalised nature of official statistics, and indeed global statistical users, demands a central focus on standards to deliver products and services that are comparable across countries. It is clear that standards must be at the core of the various modernisation programmes underway at the global level. However we are all aware that the implementation of standards in a consistent way across countries can be difficult.

There are some standards that are continuously evolving e.g. DDI. Have these standards reached a level of maturity where they can be deemed fit for purpose thus creating an opportunity for implementation across the global statistical community? Who can and will make that call for the official statistical community?

In addition, ownership and maintenance of standards relevant to official statistics can be a challenge. For example ownership of the Statistical Data and Metadata Exchange standard, SDMX, rests with the seven sponsor organisations[11] while ownership of the Data Documentation Initiative (DDI)[12] rests with the DDI alliance, within which official statistics is just one of several user groups.

---

10 http://www1.unece.org/stat/platform/display/hlgbas/High-Level+Group+for+the+Modernisation+of+Statistical+Production+and+Services

11 http://sdmx.org/?page_id=6

12 http://www.ddialliance.org/

In this context as a global statistical community how can we collaborate in an effective manner to ensure that we establish a consistent set of standards for use in modernising official statistics?

Modernisation can mean different things to different people given the variety of starting points from which compilers of Official Statistics are setting off. For example in some countries the fundamentals of an independent, objective and impartial statistical system are not enshrined in law or practice, the right of access to administrative data is not guaranteed, the data infrastructure required to exploit the vast quantity of available secondary data is not in place, the skill-sets required to engage with these new data sources are not in place etc.

The level of maturity will of course vary across organisations but also within organisations across domains or services. The concept of a Modernisation Maturity Model (MMM) is something worth considering as it could be a useful starting point from which to develop a road map towards continual modernisation.

The MMM is nothing more than a concept at this stage. However a lot of work has already been done in creating models to facilitate collaboration on an international level, GSBPM[13], GSIM[14] and most recently GAMSO[15]. The Generic Activity Model for Statistical Organizations (GAMSO) describes and defines the activities that take place within a "typical" statistical organization and builds on the GSBPM (production) to cover Strategy, Capability and Corporate Support. The development of MMM would allow the benchmarking of an NSI against a set of structured levels informed by GAMSO and give a greater understanding of that NSIs readiness for specific modernisation activities.

The adoption of an MMM would facilitate an identification of the actual modernisation challenges facing statistical organisations and support a prioritisation exercise to support investment and resource allocation. As referenced earlier in the paper it may also prove useful in identifying those areas and those institutions that are best placed to "revolutionise" and in that sense create the roadmap for others to follow when they are ready and in a position to do so (evolutionary approach).

VII.   Communication

The communications challenges facing official statisticians are many. We have a "branding" issue that needs to be addressed to maintain public trust in official statistics. This issue was mentioned earlier in the context of the "privacy-efficiency" debate.

We also have a challenge to consider in the context of the "quality-timeliness debate". In today's fast-moving and technologically advanced world where information is readily available at the touch of a screen, the value of slower and more carefully produced official statistics may become less apparent to users. However, official statistics must continue to be

---

13 http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model

14 http://www1.unece.org/stat/platform/display/gsim/Generic+Statistical+Information+Model

15 http://www1.unece.org/stat/platform/display/GAMSO/GAMSO+home

produced in a considered, professional, and methodologically sound manner, resisting the temptation to sacrifice quality for faster or more expedient products. Nonetheless the producers of official statistics do have a responsibility to work to meet the changing needs of the modern citizen and to educate users about the advantages that official statistics have over other less rigorous sources of information. This is a considerable communication challenge.

As a community we need to be clear where we stand on these key issues. Are we running the risk of being pulled into a "race to the bottom"? Are we clear about where our competitive advantage lies and what our unique selling point is?

The HLG MOS Strategic Vision attempts to address this issue of what our unique selling point is where it states that

> "The active pursuit of data and the creation of products and services that give insight from an impartial perspective, our unique selling point, will be our new mission".

Is this vision of our unique selling point shared? Are there elements missing?

Who, at the global level, is going to take the lead and develop the communications strategy and agree the "branding message"? Of course individual statistical institutes are dealing with these challenges on an ongoing basis but the delivery of a coherent and consistent message by all will have a much greater impact.

Beyond the communications challenges outlined above we also have a communications challenge related to the diverse nature of the users of official statistics. The user cohort is now much broader than before ranging from the expert researcher/academic to the citizen trying to understand the latest set of figures released.

It can be difficult to create bespoke statistical outputs for each user category while also having to create bespoke dissemination channels. But of course none of us face these challenges alone and the solutions developed by one in theory should be portable to others.

VIII.   Conclusion

One could ask how these issues relate directly to the imminent challenge related to the development of a suite of Sustainable Development Indicators (SDIs) and their subsequent measurement. This note highlights some of the challenges facing the official statistical world at a strategic level and attempts to prompt a debate on issues that will impact all statistical domains. As such, improvements that we make across the global system, should put us in a strong position to produce SDIs that are trusted, robust, relevant, independent, internationally comparable and of high quality.

There are considerable difficulties for individual statistical institutes in maintaining large work programmes while also developing new methodologies and integrating new technology and data sources. Our strength as a statistical community is in collaborating and working

together to address the broad range of issues facing us. Collaboration is the key. It's easy to say but difficult to do effectively and as a community we need to consider how best to co-ordinate the broad range of modernisation programmes that are currently underway.

In particular, there are a range of modernisation programmes underway at the international level coordinated by a range of international statistical institutions including Eurostat, UNSD, UNECE, UN ESCAP. In this context what are the most efficient mechanisms, and appropriate divisions of labour (e.g. between oversight of development and implementation), needed to support the transformative agenda?

Knowing "where" and "when" to take a revolutionary or evolutionary approach is important. This will vary by topic and institution. The MMM may be a useful tool in this regard. There is no doubt however that it is easier to make step changes (revolutionary approach) when you are following in somebody else's footsteps on a clearly defined road map. There is indeed an urgency to modernise and we must continue to use all our resources to identify the key areas for early action/intervention.

This is a high level issues paper and as such is high on questions but light on answers! Sometimes the starting point is surfacing the right questions. This paper does not purport to have raised all of the right questions, or indeed mentioned the broad range of modernisation opportunities and challenges, but hopefully at a minimum it can stimulate an interesting discussion during our plenary session in New York.