Tenth Management Seminar for the Heads of National Statistical Offices in Asia and the Pacific

7 – 8 December 2011, Chiba

MODERNISATION SPEECH

Mr. Brian Pink
Australian Statistician
Australian Bureau of Statistics

Good morning everyone and let me say it is a real pleasure to be here today as the challenges of modernising statistical systems has been of considerable interest to me for some years now. So let me start today by providing some context to these challenges.

I'm sure you are generally aware that digital information is being created and stored at a quite phenomenal rate in 21st-century societies such as our own although none of us probably appreciate the full extent of this phenomenon. Digital data is streaming in from all sorts of sources; not only from transactional and other business process, research and administrative systems of both public and private organisations, but from various sensors, instruments, on board computers, simulations and models. Enormous volumes of data are also channelling through communications, transportation, security and logistics networks. And now of course, social media such as Facebook, Twitter etc are also generating vast volumes of digital data from cameras, mobile phones and other personal digital devices. Taken in its totality, this exponential increase in digital data is overwhelming not only the capacity of many of the existing systems and tools we have designed to organise, analyse and store it but it also highlights the challenges we face in ensuring we can train the next generation of people in business, government, science and the arts with

the skills to make sense of all these data. As national statistical organisations we are not immune from these pressures.

For these almost incomprehensible volumes of data, which even a few years ago most people would not have conceived of, are, today a growing reality in many aspects of our daily lives. And although many organisations may not themselves be accumulating digital data on this scale, the challenge that nevertheless faces organisations in the 21st-century is to recognise that this digital data explosion is taking place and to work out how to take competitive advantage of the opportunities this offers them.

Such challenges are not new as illustrated by a US manufacturing firm manager who was reported to have complained about the effects of the telephone back earlier last century as follows: 'Time is lost, confusion results and money is spent'. One wonders whether the manager lasted long enough to collect the corporate gold watch. Thomas Watson of IBM of course offered the view towards the middle of last century that there would only be a need for a handful of computers in the world to meet the information processing demands of our societies.

Things have changed dramatically since then in the information, communication and computing sectors and yet it is clear that we are only really at the beginning of the ‘information age’. So let me provide you with just an example of how quickly things are changing in the data universe, and here I have drawn from an Economist article of February last year. If we look at an industry that most of us would be familiar with namely retailing, Walmart, the giant US retailer handles more than 1 million customer transactions every hour, feeding a database estimated at more than 2.5 petabytes in size (2 to the 50). And if that sounds mind boggling, Google, according to the Economist, is rumoured to process 1 petabyte of information every hour. And CISCO estimates that the

amount of data traffic flowing over the internet will reach around 670 exabytes (2 to the 60 bytes).

So What? you might ask. We've dealt with lots of data challenges to date pretty well. Yes that's generally true and we could continue to assume that what worked for us to date will continue to work for us. So let's look at that tradition.

1. Tradition – What got us through from the 80s and 90s to today

As a community of NSOs this period can be characterised by

- ☐ Siloed statistical gathering activities within our own organisations
- ☐ Much goodwill and 'opportunistic sharing' of knowledge and systems/tools
- ☐ Legacy information management systems development
- ☐ Statistical processes centred on paper forms leading to printed publications
- ☐ Initial forays into the internet, for data collection, dissemination and the use of pdfs.

2. 21st Century - A rapidly expanding set of expectations

Then over the past decade we have seen a variety of technology trends that are beginning to overwhelm the processes and systems we have developed to manage our statistical systems.

- ☐ Internet data capture
    i. Census
    ii. Business surveys
    iii. Household surveys
- ☐ Rapidly expanding data sources
    i. Administrative data – Health, Education, Taxation, etc.

ii. Transaction data – Scanner data, EFTPOS, Credit Card, etc.

For these rapidly expanding volumes of data, which even a few years ago most people would not have conceived of, are, today a growing reality in many aspects of our daily lives. And although many organisations may not themselves be accumulating digital data on this scale, the challenge that nevertheless faces organisations in the 21st-century is to recognise that this digital data explosion is taking place and to work out how to take competitive advantage of the opportunities this offers them. Such challenges are not new as illustrated by a US manufacturing firm manager who was reported to have complained about the effects of the telephone back earlier last century as follows: 'Time is lost, confusion results and money is spent'. One wonders whether the manager lasted long enough to collect the corporate gold watch. Thomas Watson of IBM of course offered the view towards the middle of last century that there would only be a need for a handful of computers in the world to meet the information processing demands of our societies.

3. Increasing demand verses decreasing resources

At the same time as NSOs are rising to meet the challenge of increased expectations our resource bases are being eroded. As we are confronted by a range of growing stakeholder expectations in areas such as

- Data integration – linking data sources, complex "information solutions"
- Internet dissemination
    i. Web services
    ii. Machine to machine connections
    iii. On-line interrogation of datasets
- Data Sharing
    i. Gov2.0

ii. Data.gov.?? websites

iii. the Semantic Web (RDF/OWL)

This digital data explosion, sometimes described as the era of ‘big data’, is one that I face in my role as head of the Australian Bureau of Statistics. Governments across Australia need sound information to inform their formulation of policies and strategies, assess the impact of their investments and refine government programs. The range of information governments require has already expanded significantly and the urgency with which it is needed has accelerated, driven by an increasingly complex and interconnected world. Some of the critical areas relate to recognition by society of urgent social problems, global financial systems failure, globalisation, terrorism, climate change and a new emphasis on sustainability. Businesses and the broader community also need reliable information for informed decision making, research and evaluation, to understand the changes taking place around them and to adjust their .

Complex problems have multiple causes, and require different information views. However, the different views need some coherence to reduce confusion and support effective evidence-based decisions. The ABS has been considering what future, new capabilities it will require to meet these new challenges, how we might work to obtain them, and in particular what actions we should be taking now. For it is my view that an effective national statistical office of the future, such as the ABS will not only need to continue to provide a framework of stable, trusted, regular and coherent key national and international statistics. In addition, we will need to be able to rapidly combine data from official statistical and a wide variety of other sources, to produce coherent information relevant to current and emerging issues of national interest. We will need to be a great deal more responsive and agile at mounting new collections to fill gaps and to answer new questions. We will maximise the availability of

information through the innovative use of statistical methods and tools which allow vast volumes of data to be stored, combined and analysed while ensuring data security and that privacy and confidentiality requirements are also preserved.

Already in my world of statistics, I see a revolution in the speed with which information is becoming available, giving rise to new information products and services and new ways of combining and using information. These alternative sources of information, some now able to be updated in real time, are available with fewer constraints and with greater detail than the information traditionally provided by the ABS. In addition to the wealth of data generated by both public and private business activity, there is an ever-increasing pool of data generated from personal devices, sensors, instruments and computers. Examples of sources are retail scanners, scientific equipment, imaging systems, transportation systems, telecommunications networks and even metrics collected about people's use of the Internet. This is resulting in a blurring of the boundaries of official statistics which is the core business of the ABS.

For example, here in Australia, there are a growing number of organisations publishing economic measures, sometimes released more frequently and often widely used to complement or pre-empt the official versions of these measures. I instance the ABS experience to demonstrate the fact that although there are great opportunities for an organisation such as the ABS to benefit from the digital data explosion, we need to be positioning ourselves now to survive and prosper in this evolving world. Unless we do this, we cannot be sure that the ABS will remain over coming decades at the heart of official information for our society. A good analogy is what is happening to many newspapers - the old business model for newspapers is being challenged by the digital revolution and many newspapers are disappearing.

4. Solutions – Joint Development & Industrialisation

- ☐ With very few exceptions, each NSO individually builds the entirety of its statistical systems.

So how might we be able to most effectively respond to all of the challenges this digital information revolution is presenting to us and should we again largely tackle these challenges independently or is there a real opportunity to work together to modernize our statistical systems.

To date most international collaboration in the area of information and communication technology (ICT) and the automation of processes and statistical methods has been between specialists. The common objectives have been to share experiences and best practices, and occasionally carry out research and demonstrate innovation. However, collaboration and the common uptake of results in this area have proven to be very difficult. For although the specialists have had the power to agree, they lacked the real authority to initiate substantial changes in their organisations.

- ☐ If we can share this development across NSO's then there are significant savings to be made - key is agreeing to standards and "sharing by design" rather than "opportunistic" sharing.

Our processes can be industrialised. It will not be in the way Henry Ford changed the world but much subtler. Mr. Ford had no really powerful tools at his disposal, only manpower. His automated factory consisted out of a long chain of people doing simple steps in the process.

Nowadays there are much more sophisticated tools available and industrialisation means a chain of very powerful tools and custom made programs that produce a product in a reproducible way. The human labour is reduced and moved to the design phase where it uses standardised means and methodologies to create a "statistical factory" for a given product.

It is in the design phase that the human knowledge accumulates.

The cornerstones of the evolving new approach to modernizing statistical infrastructure are the GSBPM, The Generic Statistical Business Process Model that is currently under administration of the UNECE, including the change management of it, together with the GSIM, the Generic Statistical Information Model, which is being developed in the Statistical Network and the CORE ESSnet project. In the practical area standardised methods and technology will enable us to rationalise. If we can realise this, we can reduce the diversity in methods, IT solutions, business concepts and information concepts thereby creating the necessary conditions to harmonise and standardise our industry.

This will allow us to convert what largely resembles cottage industries today into well engineered industrialised processes.

5. Commonality – Processes, Architecture, Standards, Technology

- ☐ Tools sharing has had some success e.g. Blaise, OECD.stat, PC-AXIS, SDMX tools, but not enough
- ☐ Still significant barriers to adoption due to adaption requirements
- ☐ We need to adopt common
    i. Processes – Generic Statistical Business Process Model (GSBPM)
    ii. Architecture – Generic Statistical Information Model (GSIM)
    iii. Standards – DDI, SDMX.
- ☐ This will provide an underlying commonality that will allow us to build and exchange applications

6. **To date there has been a number of forums where these principles have been discussed:-**

- ☐ MSIS,
- ☐ METIS,

- ☐ SAB,
- ☐ HLG-BAS, etc.

I will be discussing the outcome of a recently meeting in Geneva convened by the HLG-BAS that brought together a broad range of chairs of the various standards forums of relevance to this work in an upcoming session of our seminar.

7. ABS is interested in turning theory into practice

From an ABS perspective, we have been very keen to explore with a range of like minded NSOs, the idea of collaborating on modernising our statistical infrastructure over the next several years, sharing the gain as well as the pain.

- ☐ Initial Forays –
    i. N.Z. – Innovation in Dissemination
    ii. Canada – Web Data Collection
    iii. Norway – Editing
    iv. Sweden – Disclosure Control (Confidentiality).
    v. Australia - Operationalise a Common Metadata/Information Management Framework (OCMIMF)

Experimentation with the usage of DDI and SDMX as data/metadata formats. Lack of harnessed metadata is a major inhibiter to discovery, coherence and quality assessment as well as constraining the cost effective integration of data from different sources.

8. ESCAP – Measuring commonality of issues, interest & enthusiasm

With the constraints of limited funding, full work programs and aging infrastructure, and the challenges of a rapidly evolving environment, NSOs need to work together to succeed. Two key things NSOs need to do to speed up response to change, are to increase the relevance of statistical frameworks, standards and classifications to contemporary issues and to work out a way to more rapidly evolve, share and connect technology infrastructures. So while I

- ☐ Wouldn't begin to tell you how to run your business.

I am asking the key question and that is

- ☐ Do we all share common problems?

And if so,

- ☐ Are you interested in collaboration?

9. The Importance of Key Agreed Standards

However if we are to collaborate effectively it will be essential that we accept a far greater degree of standardization amongst us than we have been prepared to contemplate to date. And this will mean that we will have to take something from our workforce that they value very highly, some of their freedom to make intelligent decisions about how to do their individual jobs.

Still I believe it is essential to our survival to reduce the diversity **and complexity** where it cannot be justified. Moreover, the very tools and approaches that have worked well in the past may not serve us so well in the future. Much of our technology effort has been focused on supporting our individual processing models and, although we share international frameworks, information about approaches and sometimes share technology, most of our systems are individually developed for our own

purposes. Our limited technology investment resources are also being consumed by cycles of development and modernisation attempts.

If we look beyond our own statistical world, we can already observe that Many other industries and sources of data, use different metadata and data standards, but these are often common within areas of interest across the world. For example, to support banking transactions for customers everywhere, there are standards used by most banks. Major manufacturers of business equipment such as cash registers and scanners operate globally and use standards such as barcodes and RFID. This is also true for telecommunications, travel, and manufacturers of other equipment used within societies (such as traffic lights) and for spatial information (including satellite data, Google maps and GPS data).

A collective approach to determining the best ways for NSOs to incorporate this data into the information stores for their countries would be useful, as would discussion about how to judge which data might be retained to inform the future.

The use of standards and frameworks will continue to be important to support coherence in our statistics. However we also need a way to rapidly incorporate new issues and views of the data, without perturbing the base. As well as supporting international standards, NSOs have to operate in different environments with some different areas of key concern. Users of our data also need to be able to adapt or augment the standard classifications to better suit their analyses, situation, environment or to compare the data with pre-existing data provided by others. Classifications need to become more dynamic, supported by automatic ways of transforming the data quickly, such as automatic coders.

Automatic approaches are important because they enable multiple views without significantly increasing cost or time. This includes rapidly recoding existing data-sets, coding large amounts of "un-coded" data, "multi-coding" data or recoding data on demand. ABS has undertaken research into some promising new techniques, using machine learning. Because coding techniques have broad applicability and the function can be quite discrete, ABS would prefer to collaborate in the development of new coders based on these techniques. Use of common facilities in multiple organisations (including those who provide data to NSOs) would also support consistent coding and higher data quality.

I have outlined in a very general way this morning, some of the opportunities and challenges that I believe confront all NSOs as we look to define our place in the rapidly evolving information age that has exploded around us over the past decade. I hope that our discussions over the remainder of our seminar will encourage all of us to ponder some of the issues I have raised this morning and to begin to define how we might be able to share the costs of adapting our own statistical infrastructure, processes and systems to the changing world of ‘big data’. Or put another way we need position ourselves not just to survive in the digital age but to thrive.

Thank you.