Collaboration Leaders Workshop 2010  
(Sydney, Australia, 19-23 April 2010)

The Case for an International Statistical Innovation Program  
- Transforming National and International Statistics Systems

Supporting Paper
Prepared by Brian Pink, Jenine Borowik, Geoff Lee,  
Australian Bureau of Statistics, Australia

Challenges facing national statistical agencies in the 21st Century

1 The purpose of this paper is to identify the type of organisations we need to become. It is time to consider how to evolve as National Statistics institutions for the future. With the constraints of limited funding, full work programs and aging infrastructure, and the challenges of a rapidly evolving environment, we need to work together to succeed. Two key things we need to do to speed up our response to change are to increase the relevance of our frameworks, standards and classifications to contemporary issues and to work out a way to more rapidly evolve, share and connect our technology infrastructures.

2 National Statistical Institutions (NSIs) face many critical challenges in the 21st Century. Governments and international organisations around the world make large investments to manage critical issues. They need information to help the formulation of good policy, assess the impact of the investments and refine government programs. The range of information required has 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.

3 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. NSIs have a key role in assisting coherent and logical analysis of data
from multiple sources. The fundamental business of national statistics institutions is to assist
governments, business and communities to make informed decisions. As institutions,
individually and collectively, we work to provide a coherent base of information with accurate
measurement of change over time in key economic, social, demographic and environmental
indicators. We provide additional value through our attention to coherence in a number of
dimensions: particular areas of our nations; key issues and groups (the young, the aged,
indigenous people, disadvantage, health, the economy, the environment, etc); comparison of our
individual nations with similar nations; and international statistics (either for regions of the world
or the entire world).

4 Citizens are expecting more from government. There is a resurgence of social conscience
in particular fields, perhaps enabled by improved education and access to information and like-
minded people through technology. There is often strong community interest in assessing the
success of government in tackling problems and in working with government to design effective
programs. The mechanisms for the community to raise issues with government are also
expanding in range and power with the increasing use of the internet to circumvent traditional
media and public relations mechanisms. The debates are becoming more open, visible and
volatile. Governments are increasingly looking at how to use web 2.0 and the emerging web 3.0
to transform government services and interaction.

5 The revolution in the speed with which information is becoming available is creating new
information products and new ways of combining and using information. These alternative
sources of information are sometimes updated in real time and are available with fewer
constraints and with greater detail than the information provided by NSIs. In addition to the
wealth of administrative data generated by 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, transport systems, telecommunications
networks and even metrics collected about people's use of the Internet. There is a blurring of the
boundaries of official statistics. A number of organisations are publishing statistics, sometimes
released more frequently and often widely used to complement or pre-empt the official versions.

6 The key part about the internet and the knowledge economy is not where is it now, but
where will it be in 10 or 20 years time, and what NSIs need to do to survive and prosper in this
evolving world. Unless we do this, we cannot be sure that NSIs will remain at the heart of official information for societies. A good analogy is what is happening to newspapers - the old business model for newspapers no longer works and many newspapers are disappearing.

7 Governments seem unwilling to provide the additional investment necessary not only to maintain but expand the core existing authoritative baseline data, but are often willing to invest in rapid responsive approaches which provide answers to new challenges and questions they are confronting.

8 The current approach to developing international and national statistical frameworks, standards and classifications is slow and considered. The approach suits the need for comparable statistics over time, but is a significant barrier to rapid change. It can take decades and great expense for new versions of existing standards to be implemented across the world, and for many pragmatic reasons some countries do not ever implement them. This means that often our current data are not seen as fully relevant to new areas of information enquiry, particularly when compared to some of the more agile and informal information sources. The outdated nature of our frameworks can also complicate the analysis task, as users attempt to make their own adaptations to suit the real world. We need to consider approaches which preserve the comparability of statistics but enable rapid ways of producing additional statistical models to reflect contemporary changes taking place in our economy and society and in our different national interests, international interests, contemporary activities and issues.

9 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 resources are consumed by cycles of development and modernisation attempts.

A Vision for 21st Century NSIs

10 NSIs collectively have an important role in evidence-based policy making through enabling comparison of data across regions, countries and the world. By supporting international statistical standards, we provide governments with the opportunity to assess the success of similar
policy interventions in other countries, whether prior to making significant investment, during implementation or once programs are established. For global issues, the importance of coherence across the international statistical system is even more pronounced.

11 The effective NSIs of the future will continue to provide a framework of stable, trusted, regular and coherent base of key national and international statistics. In addition, we will be able to rapidly combine data from official statistics and a wide variety of other sources to produce coherent information relevant to current and emerging issues. We will be 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 methods and tools which allow data to be combined and analysed while ensuring privacy and confidentiality requirements are preserved. We will further develop and share our strong information management capabilities and work effectively with others to encourage innovation in information management and use.

12 As new information sources are developed or discovered we will incorporate the most useful into the enduring fabric of information available to our governments for informed decision making and into the base of important international statistics. NSIs could also take steps to shape how new information sources are developed and used. We have the tools, the people and the credibility to be a creative partner in this field.

13 The ABS has been considering what future, new capabilities might be required to meet these challenges, how we might work together to obtain them, and what actions we should take now.

**Future Capabilities**

*From static data products to "common information services"*

14 Typical NSI websites contain a mix of information products. These include electronic publications supported by more detailed data products in a variety of formats (spreadsheets, data cubes, data warehouses) and other products such as confidentialised microdata files and information about standards and classifications. The standards and classifications are often not explicitly linked to the data products which use them and searching and finding specific data is often a manual process, unique to each organisation's website.
15 In the future it should be possible to bring together information automatically and repeatedly, in ways we cannot imagine and at speeds we can't envisage. We need to support the demand for assembly of data from different sources, including our own, and at the same time protect the confidentiality of the data. We have, individually and collectively, significant data assets of value to others. Society and governments are exploring ways to unlock public sector information and a shared approach by NSIs has the potential to assist this drive. We should consider how to establish common information services to attract development investment from the commercial, government and non-government arenas to support this future. If enough momentum is achieved through a common approach, other organisations and individuals will chose to connect their services and innovations to ours. By working as a group this will not only make it easier for these organisations to deal with us, but should result in richer services and reduced costs.

16 To be successful, with a "common information services" business model:

- NSIs would have common metadata services on their websites. The metadata services would provide descriptions of the data elements in each information product including definitions of the rows and columns of each table or file, links between data element definitions and data products, links to data quality information, descriptions and concepts. Visitors to NSI websites would be able to browse or search the repositories to find information about which data elements are available, in which combinations, and from which specific data products. It is important that the information is also machine searchable and provides metadata services to other applications both on NSI websites and elsewhere, allowing clients to rapidly assemble data from multiple sources.

- As a companion to a common metadata service, NSIs would implement one or more common data services. We are already struggling to produce and quality assure the many different cross tabulations that users require and the process of individually crafting these takes resources and time, slowing down our ability to respond quickly enough. Many NSIs have recognised that the solution is to invest in dynamic data services which allow clients to specify the information they require and generate this dynamically from 'raw' data. Examples of this already exist (for example, the ABS Census TableBuilder product).
• As well, NSIs would have common confidentiality services. Because of the need to protect confidentiality, NSIs typically hold a great deal more information than they can release. There are complex approaches for ensuring that data released meets confidentiality requirements. For example, the Census TableBuilder product applies a perturbation method which 'works' for simple count data but is not directly applicable to other types of data. Other types of data might include continuous variables, longitudinal data or composite data sets. There is a need for several different confidentiality methods, depending upon the characteristics of the data being accessed.

• Integrating common data, metadata and confidentiality services, we also need to design approaches which support the ability to dynamically link our data and the data and services from elsewhere. Dynamic data linking services offered by our organisations and though our websites should allow comparison and combination of data from multiple sources.

• NSIs would partner with industry and with government to stimulate and enable new ways and means for individuals, businesses and communities to gain access to and to use publicly available data.

17 The combination of these services would also provide important functionality for future access to more sensitive microdata. ABS' current Remote Access Data Laboratories (RADL) provides a batch submission process, which allows researchers to submit programs against data that is not available to them in raw form. It has a heavy reliance on assessing and pre-confidentialising the unit data being analysed. Future development aspirations are to move this towards an interactive service, with the focus on dynamically confidentialising outputs before it is provided to the client. This is far from a trivial challenge. However, NSI expertise in methods for "confidentialisation" of data is relevant to all data custodians. Further development of common approaches could be used in many organisations to support information sharing, to increase access to statistical data and potentially to support new data linking approaches across multiple organisations’ data.

18 There will be many very important cases where the information required by clients will still be built from pre-assembled data products. Some obvious examples are time series of data, products such as the national accounts where the information is assembled from multiple sources
within a standard framework, products like price indices where the assembly process involves complex calculations such as chain linking, and very large datasets where the processing cost of creating everything dynamically might be too high. We should also share investigations into approaches for managing these cases and providing them to our clients through the same metadata and data services mentioned above.

From publications to communication

19 Publications from specific data collections will remain important, but there will be a subtle change in their role. They will become alerts, which warn data users that new information has become available, that announce the key headline stories in the data in a way that other agents, including the media, can easily convey to the community, and that direct clients via active links to the richer data available through the data and metadata services.

20 This separation of the data services from the presentation process will support the use of new ways of exposing and communicating the meaning stored in statistical data, such as those described by Professor Hans Rosling, Professor of International Health at Karolinska Institute and Director of the Gapminder Foundation, which developed the Trendalyzer software system. There are many other examples of innovation in this area (for example, mashupaustralia.org). The techniques available will continue to evolve. Successful NSIs in the future will provide their information in ways which support and encourage these different ways of communicating the underlying messages in the data.

Support for transaction data flowing at a much higher volume

21 The volume and sources of real-time or near real-time data are increasing exponentially. Different industries and sources 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 many 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 NSIs 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 future
policy issues. This work could include determining common approaches for data discovery, data exchange and, perhaps, data warehousing/archiving and access to data.

Ability to rapidly incorporate new issues and views of data into standards and classifications

22 The use of standards and frameworks will continue to be important to support coherence. However we need a way to rapidly incorporate new issues and views of the data, without perturbing the base. As well as supporting international standards, NSIs have to operate in different environments with some different areas of key concern. Users of our data also need to adjust 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 be dynamic and 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 NSIs) would also support consistent coding and higher data quality,

'Rapid-response' capability

23 Often existing collections of data are not sufficient to answer new questions. If statistical planning processes are operating well, we should have the basic baseline data to shed light on an underlying social, economic, or environmental issue. However, when a policy direction is set, and an "experiment" is underway, governments will need up to date information about the outcomes and how these relate to their investment. They will need to make comparisons with what has happened in the past, or is happening elsewhere.

24 Successful NSIs will be required to mount new collections, or adapt existing ones, quickly and responsively. These new collections may be 'traditional' data collections run by an NSI, or they may involve diving into data which is available through other organisations' systems or websites, or both. Culturally, NSIs have a strong bias towards stable data series, and developing collections using careful evaluation of a wide range of user needs, thought into the
underlying conceptual framework, careful development and testing of questions and
questionnaires, samples designed for optimal efficiency, processing systems tailored and tested in
advance, data edited comprehensively, information analysed carefully and publications
constructed, presented, etc. All these are valuable approaches, but additional and different
approaches will also be crucial.

25 Technology or methodological changes which will help us to achieve more rapid
response, but changes to the mind set, the capability and the capacity of skilled staff are also
fundamental. We need to be able to form agile teams with a combination of relevant expertise
(statistical, analytical, technical and policy related) and equip them with the mandate and tools
required to respond quickly. In addition to the capabilities already mentioned, some of the tools
and infrastructure required are question modules, web survey facilities, call centres for follow-up
or interview, ETL (extract, transform, load) tools, additional confidentiality approaches, data
warehouses to store the results and web site facilities to share them. We need agile legal
approaches that can help to remove barriers to data exchange such as template data sharing
agreements and licences such as Creative Commons. We need engagement mechanisms which
make it easy to involve people (with academic, commercial, NSI or other backgrounds) in the
work of these teams and if possible shared approaches to the development of these people.

Connecting processes and passing metadata and data easily between them

26 Once an initial rapid response has proved its worth, provided that the underlying demand
remains, successful NSIs will have to improve and institutionalise the new information stream
into their existing core production processes.

27 A commitment to a standard way of describing our information using a format
(SDMX/DDI is proposed) would allow us to connect our statistical process steps together more
easily over time as we develop or redevelop key parts of our systems. It would also enable others
to provide us with functionality more easily.

Analysing assemblies of data

28 Tools being developed in the Web 2.0 / semantic web space hold out the promise of new
and more efficient ways of analysing metadata stores. The use of a standard approach such as
SDMX/DDI containing links between particular aspects of information (eg. data collections
which use a specific data element, questionnaires and forms used to collect the information, data
quality statements about the collection) will support improved and better directed investigations by skilled statistical staff into any discrepancies between disparate data sources. It will also provide insights into how the data should or should not be used.

**How We Might Work Together**

29 The areas mentioned above are examples of where we might profitably collaborate and innovate as a community of NSIs. Each NSI is likely to have a view of where investment in change is required and the development of a program for collaboration on innovation would require us to explore these views and agree on some initial candidate areas.

30 As individual NSIs, our current approaches are not as adaptive as we would like. Our strategies have been to either attempt large "whole of system" transformation programs, or to deal with particular parts of the process at particular points in time. We have tended to engineer our processes and systems for current problems rather than looking at what we need for the future. This strategy is only appropriate for a relatively "steady state" environment. The level of change required is profound and it is unlikely that any NSI will be able to make the investments required to facilitate change at the rate required on its own. The technology environments that are in use today are likely to be totally irrelevant in the future, so our current approach of redeveloping large components of our infrastructure every 5-20 years will not serve us well. Instead, we need an approach that allows us to evolve smaller components of our statistical infrastructure much more rapidly, and which allows us to inject and attach new functionality as it appears in the environment.

31 We need to encourage innovation in our own organisations but also excite and use interest in statistical innovation in academia, other information providers, the ICT industry and in creative individuals, so that we don't have to fund all of our change directly ourselves. We need the ability to easily combine these innovations along with other tools from the knowledge economy, as they emerge. As a group, we should be able to spark interest amongst other NSIs, in research and academic communities and in the ICT vendor community. Researchers, academics and students probably have the time, energy and interest to innovate and ICT vendors tend to move into areas that address common issues. There are many groups of dedicated developers/enthusiasts who like to innovate in their area of interest and who are connected via the web, including groups GGobi who like to develop new ways of analysing data and numerous
groups of open source developers with specific interests (see sourceforge.net). These enthusiasts are a vast untapped workforce. In the commercial arena, Apple has published the details of how to write small applications for the iPhone/iPod and allows developers to place their applications in an on-line store. Every time a particular application is downloaded, the developer receives a small royalty fee. In the short time since the iPhone was released, many thousands of applications have been written. Most are trivial, but there are increasing numbers of useful applications becoming available. Use metrics are collected so that royalties can be paid and as a consequence, the popular applications are well known. Apple have stated that there are now more than 85,000 applications available in the App Store and more than 2 billion iPhone and iPod Touch applications have now been downloaded. Although these examples may not work for NSIs, we need to explore ways to stimulate contributions from a wider community.

32 As an international statistics producing community, we need to try to build standard approaches together while we still have the relevance to force a standard. It is important to make some decisions as a community about what needs to be standard, choose some standards even if they do not represent the best "technical" approach and then evolve these over time if required. As a community of organisations our "buying power" can be amplified if we can agree on common requirements and design. It may be difficult to reach agreement on the requirements for large components (such as an entire input data warehouse), but if we start with components for some of the smaller, relatively discrete functions of statistical processing, these could be built into larger components over time.

33 To be effective, the program needs to support change through processes, methodology, technology and people. The people dimension could include opportunities for sharing staff across NSIs (either through short or medium term secondments or through the use of virtual teams). It is likely that the skills, capability and careers of staff in our organisations would be enhanced through the program, particularly if we can include collaboration on the development and delivery of relevant training. Many NSIs also have active international statistical aid programs and there is potential to link these to an innovation program, with benefits for both programs.

34 Although there are significant institutional and other barriers, we have been encouraged by the success of a number of open source and cross government projects. More effective
collaboration has the potential to reduce the costs for each statistical organisation of building and maintaining our software and could allow us to make stronger investment in particular areas of our work. There is also the potential to accelerate the availability of statistical infrastructure above what we could each achieve individually and at the same time to enhance support for the statistical systems of developing nations. Combined with efforts to enhance statistical skills, a stronger base of statistical infrastructure shared with other organisations should assist our statistical leadership aspirations, reduce development and cost duplication, and could ultimately drive the emergence of a truly integrated global statistical information system.

35 International organisations, particularly the UN Economic Commission for Europe (UNECE), EUROSTAT and OECD, have driven frameworks such as the Generic Statistical Business Process Model (www.unece.org/stats/gsbpm) and the Common Metadata Framework, and recommended metadata standards such as DDI and SDMX through active engagement with NSIs. These are a very useful starting point providing a top-down view of our business and important guidance regarding the ways to link processes and components. So that components will integrate, there is a strong need for a standard way for the components to share data and metadata. A number of NSIs and international statistics organisations support the use of DDI and SDMX for this purpose. Part C of the Common Metadata Framework http://www.unece.org/stats/cmf/ provides further information.

36 What is now needed is a practical way to work together to address some of the challenges outlined above.

37 It is recommended that a small group of National Statisticians be invited to explore the potential for establishing a joint collaboration and innovation program. Areas for discussion and decision include program outcomes, principles, governance, selecting potential projects, investment models and funding contribution, personnel exchange and management. If agreement can be reached by this group, it is expected that a working group would be formed to establish and implement the program.

38 The following discussion outlines a possible some aspects of an approach.

39 Many NSIs have endorsed the Generic Statistical Business Process Model as a reference model. This model provides a useful description of our business in a way that could assist discussions about the development of shared approaches. It is also useful within NSIs for
discussing the areas ("hot spots") where investment is most required (the red boxes in the diagram are examples of current interest to the ABS).

Figure 1: Generic Statistical Business Process Model (top level) showing some areas of current ABS interest
Rather than commencing a huge high risk project to develop a single solution covering the entire statistical business process model, an incremental approach is recommended. An approach is outlined in the diagram below.

**Figure 2: An Incremental Business Transformation Approach**

A small set of projects should be established to develop the first group of common components or approaches. A set of criteria should be established to assist this selection but would might include attributes such as: contributes to 'future proofing' of NSIs; required as early 'building block' for future collaboration; commits to a required standard; delivers tangible benefits not only to individual NSIs but to the broader statistical community; usefulness to more than one participant; ability to complete in defined time period; ability to incorporate easily into existing approaches, speed of return on investment, etc.

Each project will need to progress through a consistent process of determining shared requirements, preparing a project proposal, common design, cost and benefit analysis, team establishment (virtual or co-located), project plan and milestones for delivery. Note that participants may not contribute directly to all of the projects.
43 It will be important to select an initial set of projects that have a good chance of success, as it is envisaged that each year the group of National Statisticians, acting as 'venture capitalists', will on the basis of results, assess whether to continue and expand the program and which projects to support.

44 It is hoped that as the program gains momentum, there will be many other potential collaborators for this work. These include other government, non-government and commercial organisations within our nations, the full range of NSIs, international statistical organisations, academia, global ICT organisations and creative individuals. All of these are likely to have capabilities, components or plans that might be useful. The development of components of the statistical process by one NSI which are then shared with others sometimes works, but in the past this has tended to be opportunistic rather than planned. There will be many solutions that fit into each box of the GSBPM and this is good - competition between NSIs, commercial providers and others will only serve to accelerate the change and innovation required.