

Unclassified

STD/DOC(2003)3



Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

18-Apr-2003

English - Or. English

STATISTICS DIRECTORATE

STD/DOC(2003)3
Unclassified

STATISTICAL DEVELOPMENTS AND STRATEGIES IN THE CONTEXT OF E-GOVERNMENT

Brian Finn and Enrico Giovannini

Contacts:

Mr. Brian Finn, tel. (+33 1) 4524 1309; Brian.Finn@oecd.org

Mr. Enrico Giovannini, tel. (+33 1) 4524 1911; Enrico.Giovannini@oecd.org

JT00143051

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

English - Or. English

The OECD Statistics Working Paper Series – managed by the OECD Statistics Directorate – is designed to make available in a timely fashion and to a wider readership selected studies prepared by staff in the Secretariat or by outside consultants working on OECD projects. The papers included are of a technical, methodological or statistical policy nature and relate to statistical work relevant to the organisation. The Working Papers are generally available only in their original language – English or French – with a summary in the other.

Comments on the papers are welcome and should be communicated to the authors or to the OECD Statistics Directorate, 2, rue André Pascal, 75775 Paris Cedex 16, France.

The opinions expressed in these papers are the sole responsibility of the author(s) and do not necessarily reflect those of the OECD or of the governments of its Member countries.

<http://www.oecd.org/std/research>

Copyright OECD, 2003

Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publication Service, OECD, 2, rue André Pascal, 75775 Paris, Cedex 16, France.

Abstract: This working paper has been written as a contribution to the OECD e-government project launched in 2001, which explores how governments can best exploit information and communication technologies (ICT) to enhance good governance principles and achieve public policy goals. The paper highlights the way in which the collection, compilation and dissemination of statistics has changed dramatically as NSOs have taken advantage of the opportunities afforded by ICT advances. In this regard, it describes changes that have occurred in national statistical offices (NSOs) in response to growing citizen demand and outlines both developments that have been made possible and necessary by recent technological advances in software, communications and computing. In addition, the paper shows that statistical institutions have a significant role to play in e-government developments, having often been given a major role in national e-government initiatives. Finally, it shows that the drive towards e-government requires fundamental changes in the way that public services are delivered, including cultural changes within public administrations, and changes to operating practices such as better interaction within and between public agencies.

Résumé : Ce document de travail constitue une contribution au projet de l'OCDE sur le gouvernement électronique, lancé en 2001 et explore les moyens pour les gouvernements d'exploiter les technologies de l'information et des communications (TIC) afin d'améliorer les principes de bonne gouvernance et de réaliser les objectifs des politiques publiques. Ce rapport met en lumière la manière dont la collecte, la compilation et la dissémination des statistiques ont nettement évolué au fur et à mesure que les offices nationaux des statistiques ont tiré profit des opportunités permises par l'avancé des TIC. À cet égard, ce rapport décrit les changements qui se sont produits au sein des offices nationaux des statistiques en réponse à une demande croissante des citoyens et souligne tous les développements rendus possibles et nécessaires par les progrès technologiques récents dans les domaines des logiciels, des communications et de la programmation. En outre, ce rapport démontre le rôle significatif des offices nationaux des statistiques dans les développements de la mise en place d'un gouvernement électronique, après avoir souvent été sollicité dans le cadre des initiatives nationales de gouvernement électronique. En conclusion, le rapport démontre que le chemin vers le gouvernement électronique exige des changements fondamentaux dans la façon dont les services publics sont rendus, y compris des changements culturels au sein des administrations publiques et des changements dans les procédés de fonctionnement, tels qu'une meilleure interaction entre les agences publiques ainsi qu'à l'intérieur de celles-ci.

TABLE OF CONTENTS

STATISTICAL DEVELOPMENTS AND STRATEGIES IN THE CONTEXT OF E-GOVERNMENT....		5
1.	Introduction	5
2.	The creation of a “national knowledge base”	8
2.1	The changing role of statistical institutions in responding to society demand for information ..	8
2.2.	Developing national gateways to official statistics.....	9
2.3.	Towards free dissemination of official statistics.....	13
3.	Statistical languages and models for the exchange of information	16
3.1	Role of metadata in statistics	16
3.2	New standards and statistical languages for exchange of data and metadata	17
3.3.	SDMX initiative.....	18
4.	Statistics and confidentiality.....	20
4.1	The key issue of confidentiality for statistics.....	20
4.2	Legal and administrative measures for confidentiality	22
4.3	Methodological and technical measures for confidentiality	24
4.4	Innovation and regard for confidentiality	25
5.	Conclusions	28

STATISTICAL DEVELOPMENTS AND STRATEGIES IN THE CONTEXT OF E-GOVERNMENT¹

1. Introduction

This paper deals with recent developments in statistics in the context of e-government. It has particular regard to the perspectives of the OECD e-government project launched in 2001,² which explores how governments can best exploit information and communication technologies (ICT) to enhance good governance principles and achieve public policy goals. That project considers the impact of e-government on the structure, operations and capacity of public administrations from a number of different perspectives such as:

- What are the changing characteristics and demands of society and how can the greater use of ICT within government facilitate the capacities of administrations to respond?
- What progress has been made so far and what are the good practice strategies and solutions to problems that have emerged?
- What needs to be planned for and done to achieve the desired longer-term outcomes?

With regard to these perspectives, this paper describes changes that have occurred in national statistical offices (NSOs) in response to growing citizen demand and it outlines both developments that have made possible, and developments that have been made necessary, by recent technological advances in software, communications and computing. It highlights the way in which the collection, compilation and dissemination of statistics has changed dramatically as NSOs have taken advantage of the opportunities afforded by ICT advances. Indeed, while it is not universally recognised, the changes in work practices have been so fundamental that NSOs are usually to the fore of the e-government process in their countries, whether through providing static information only, interactive services for data manipulation, or sharing information with international agencies such as the OECD or the International Monetary Fund.³

While many of these changes have been made technically possible by the ICT developments of recent decades, the rationale for change has been driven by user demands for more information, disseminated on a more cost-effective basis. This paper highlights how statistical institutions – both

1 This paper in part draws on work relating to statistics and e-government previously done in the OECD by Ms. Ambra de Baysar of the Statistics Directorate, which the authors gratefully acknowledge.

2 The project is being undertaken by the Public Governance and Territorial Development Directorate of the OECD, with inputs from other Directorates in areas such as e-commerce, information economy, statistics, etc. It will produce research papers, policy briefs and specific reports with a flagship report to be finalised in May 2003. There will be a major conference to consider the project's findings following the publication of the report. At that stage, it is envisaged that a second two-year phase of the project will begin.

3 In the United Kingdom Government publication *Information Age Government: Benchmarking Electronic Service Delivery*, reference was made to three stages of development:

Stage 1 - web site presence with on-line information and forms

Stage 2 - interactive services with ability to search in databases and complete forms on-line

Stage 3 - integrated services with which agencies can share authenticated data with other agencies.

NSOs provide services at all these stages (albeit at different levels of sophistication).

national and international – are responding to the challenges of providing statistics in a manner that satisfies customer needs in a more cost-efficient manner. As citizen demand or need is at the core of most e-government policy initiatives, the paper deals therefore with a topic that is relevant to government agencies in general. Thus, although the direct focus of this paper is on statistical developments, it should be of interest to anybody concerned with implementing or involved with an e-government process.

The term “e-government” tends to focus on the use of new ICT by governments as applied to the full range of government functions. Of course, the networking opportunities offered by the Internet and related technologies have begun to transform the administrative operations of government by:

- enabling citizens to access government information more readily;
- facilitating general compliance (licenses renewals, building permits, etc.);
- helping citizens in having access to personal benefits and in paying taxes;
- developing e-commerce dimensions between administrations (government-to-government or G2G) and with businesses (government-to-business or G2B);
- integrating of government-to-government information and services;
- stimulating a more active participation of citizens in decision-making processes (discussion groups, voter registration, on-line voting, etc.).

However, the overall objective of most e-government policies is to make government more accessible to citizens and to improve efficiency by helping other parts of society (enterprises, households and private institutions) to minimise transaction costs in dealing with administrative matters. In this regard, ICT is just a tool to facilitate the realisation of this objective. E-government should be about taking advantage of the opportunities presented by ICT developments to deliver higher quality public services to a wider community. It is important not to lose sight of the fact that e-government should be more about government than about “e”.

E-government initiatives can be interpreted as part of the wider development of the so-called “information society”, i.e. the society where “people, enterprises and governments want to make decisions and assessments, and reach judgements on the basis of objective data, specific data, on numbers with a sound basis”.⁴ They provide an opportunity to develop a new relationship between governments and citizens based on greater trust and citizen participation. On the other hand, the implementation of e-government policies raises important governance issues and concerns about protecting citizens’ expectations of privacy and data security. For example,

- governments need to develop very precise criteria for the release of public information that may contain personal or sensitive data;
- difficulties in evaluating the final outcome or effectiveness of e-government policies are emerging;
- the role of “digital divide” in widening and deepening existing forms of social exclusion needs to be carefully evaluated;
- more generally, there is a risk that lack of analysis and planning could cause the failure of e-government projects.

4 J.Nuno Mendes in “Information and Knowledge: The Role of Statistics”, *Proceedings of the 86th DGINS Conference*, Porto, June 2000.

In the development and management of information societies, statistics play a relevant role in many respects. First of all, statistics basically respond to the information needs of general government, economic agents and citizens in assessing the state and the evolution of economic, social and environmental phenomena at international, national and local level. Secondly, statistical methods are at the core of quantitative analyses that help to identify best policies or individual decisions. Finally, statistical information is a fundamental tool in democratic societies to monitor policies and resource allocation. It maximises the capacity of people to objectively assess the efficiency and effectiveness of policies and to elect their representatives having regard to the maximum amount of available information. Thus, if linkage can be developed between statistics and e-government initiatives, some of the problems associated with such initiatives can be reduced.

Coherent linkage between statistics and e-government policies has been established in several OECD countries. In particular, five main roles for statistics in supporting and orienting e-government strategies can be identified:

- statistical products (publication, databases, etc.) contribute to the creation of a “national knowledge base”, in the context of the information society;
- statistical languages and metadata models⁵ contribute to facilitating the exchange of information across different parts of the public administration;
- statistical offices are increasingly extracting useful information from administrative data and maximising the knowledge of economic and social processes, without imposing an excessive burden on respondents;
- statistical techniques for protecting the confidentiality of respondents (citizens, businesses, etc.) contribute to minimising the risk of abusing the privacy of individuals associated with the collection, exchange and dissemination of information for administrative purposes;
- statistical surveys and models contribute to the objective identification of citizen need and statistical indicators can be developed to evaluate (ex-ante and ex-post) the efficiency and effectiveness of public policies.

As this paper deals with statistical authorities meeting the challenge of using ICT advances to provide more cost-effective services, its focus is primarily on the first four points above. However, this does not imply that nothing is being done with regard to using statistical data for the evaluation of e-government policy initiatives. For example, the Australian Bureau of Statistics (the ABS) has published statistics on the extent of computer and Internet access in Australia.⁶ Furthermore, as part of its work on the Information Economy, the Science, Technology and Industry Directorate of the OECD has published

5 Statistical metadata describe statistical data and – to some extent – processes and tools involved in the production and usage of statistical data. Structural metadata refers to the concepts, structure definitions and code lists defined by a centre institution for the exchange of statistical information with its partners and other organizations. Metadata models include:

- A metadata content standard, which specifies a set of metadata items and does not specify the physical format of the content, the services to be provided or the syntax used
- A metadata registry, which is a system for managing structured metadata describing the semantic content of shareable data and metadata
- A statistical metadata repository (MDR), which is a logically central statistical metadata repository that allows for the query, editing, and managing of metadata. Such a system provides a mechanism for looking up information about statistical products as well as their design, development, and analysis.

6 For details see: <http://www.abs.gov.au/Ausstats/abs%40.nsf/e8ae5488b598839cca25682000131612/ae8e67619446db22ca2568a9001393f8!OpenDocument>.

Measuring the Information Economy. This contains over 80 indicators based on the most up-to-date official statistics in Member countries.⁷

One of the most important e-government initiatives of recent years is the European Union's *e-Europe*, which aims to accelerate Europe's progress towards the digital age, in a socially inclusive manner.⁸ In the context of using statistics to evaluate effectiveness of public policies, it is interesting to note that *e-Europe* makes extensive use of benchmarking to assess whether the initiative is meeting its goals. It does this in a number of key areas, including accessibility to ICT and Internet costs. These are of course factors in overcoming the problem of digital divide, where the very people who have most to gain from e-government cannot use the service either because they cannot afford to, do not how to, etc. If citizens are to take advantage of the better quality services available, they must be able to access them. This includes statistical information disseminated by NSOs and other statistical agencies – if citizens cannot use the better quality information being published, the improvements are ineffective. Nevertheless, the issue of developing capability to use the Internet is a matter of government policy and as such is not directly within the competence of statistical authorities.

2. The creation of a “national knowledge base”

2.1 *The changing role of statistical institutions in responding to society demand for information*

In the “information society era”, information is, by definition, the critical factor for the development of society and the competitiveness of an economic system. ICT facilitates the rapid exchange of information among citizens and businesses and the Internet has produced a real “cultural revolution”. Large and small on-line press agencies, web-sites created by businesses and associations, portals to access public administration bodies, etc. provide today's citizens and businesses with a mass of information to use in making their decisions.

The use of statistical data in public decision making processes has become more evident over the last decade in all countries. For example, in the preparations for European Monetary Union, statistics were fundamental in evaluating individual country performance compared to the famous “Maastricht criteria”. This helped to raise awareness of the importance of statistics in decision-making, which in turn increased demand. Nowadays, both businesses and citizens ask for more information to take decisions or to evaluate the performance of governments, while newspapers and media in general now pay more attention to both short-term economic indicators and structural statistics.

This interest in statistical data is not confined to that provided by national statistical authorities. For example, specialised research institutes and companies collect and disseminate data on specific areas (market trends, large enterprise balance sheets, agents' expectations, etc.) and business associations have begun to manage and disseminate statistical information to their members and associates, merging different elements, making more estimates, etc. As a consequence, national “public” data providers (NSOs and other government bodies) and international organisations have found themselves, in terms of quality, competing in the information market with other data providers, many of which seem able to provide data on a more timely basis than public providers. While this timeliness may be at the expense of other important factors, many users are unaware of or unconcerned with the various quality dimensions of statistics. For example,

7 For details see <http://www.oecd.org/EN/home/0,,EN-home-110-1--1-no-0,00.html>.

8 For details see http://europa.eu.int/information_society/eeurope/index_en.htm.

timeliness and accessibility for economic statistics are often given perhaps undue preference to other dimensions of data quality (such as accuracy, comparability, coherence and interpretability).⁹

However, the advent of the Internet, the improved performance of servers and PCs, and the development of software capable of managing very large databases, has enabled national statistical authorities to disseminate statistics more rapidly. The issue of timeliness has been addressed by all statistical bodies and has resulted in an impressive improvement in several fields. The issue of accessibility of statistics has also been investigated in the context of competition with other providers of data, both official and non-official, and from the point of view of the final user. The availability of powerful servers and PCs has provided new opportunities to place large databases on the Internet. New software has been developed to create and manage very large databases, containing millions of elementary data or time series. These databases are accessible through the Internet and data can be extracted using different formats, according to user preference. As a result of these developments, public data providers are beginning to match the performance of their private sector counterparts and can offer tailored products, updated in real time, using a wide range of statistical tools to elaborate elementary data and to release “ready-to-use” products.

From a technical perspective, these developments have completely changed the previous technical and organisational framework. All statistical data providers today consider the Internet to be the most important instrument for disseminating statistics. This platform really comprises two elements: data and statistical methodological information (metadata) outlining national practices. Guidelines have been proposed for the dissemination of data and metadata through the Internet and new languages have been adopted or are under development to ensure the best use of this dissemination medium.¹⁰

In conclusion, the development of the information society has made it more difficult to maintain the usual distinction between “public” and “private” (or between “official” and “unofficial”) statistical data and has obliged public bodies to adapt their strategies to the new environment. Moreover, e-government strategies push public administrations to provide as much information (including statistics) as possible to the public, although this effort can increase user confusion in terms of duplication of web sites and statistical data. This risk is particularly high in countries where the national statistical system is decentralised, i.e. where single government departments or regional statistical agencies regularly produce statistical information about their own field of competence on their own websites. This problem is considered in Part 2.2.

2.2. *Developing national gateways to official statistics*

To overcome the dispersion and diversity of multiple websites, many governments have a specific objective of providing access to all government services through a single portal. A portal is a “one-stop shop” that provides a wide range of information and services from a single entry point. Such a facility has the advantages of offering citizens one point of access with no need to log on to multiple websites, and encouraging a more standardised approach by government agencies that often do not share a common navigation system or presentation style. So, although portals require significant back-office integration and co-ordination, they can act as a catalyst for ensuring greater consistency between agencies. In Canada, for example, the “Government On-line” project was launched with the aim of putting all public services on-

⁹ For this reason, international organisations and statistical offices have undertaken initiatives to establish guidelines and minimum standards in disseminating statistics (see, for example, the IMF initiative on Special Data Dissemination Standard (SDDS) at <http://dsbb.imf.org/sddsindex.htm>).

¹⁰ For more detail on this issue, see Joint UNECE/Eurostat Work Session on Statistical Metadata, Working Paper No. 9, “The Structure of the Metadata System at Statistics Sweden”, Statistics Sweden, March 2002.

line by 2004. The vision of the project is continually to improve the quality of interaction between Canadians and their government by enabling citizens to request and receive services and information when and where it is most convenient for them, wherever they live. The information is provided through one-stop access points with services and information organised by theme or type of activity, rather than by government department.

To provide users with statistical information produced by public bodies in a co-ordinated way, NSOs in several OECD countries have been asked to lead national (or federal) projects in the framework of e-government policies. This is quite logical given that NSOs have considerable experience in disseminating data on a wide variety of subjects in a meaningful way to users. In Canada, "Statistics Canada is a key player and content provider to the government's portal, Canada Site www.canada.gc.ca. To ensure more seamless access to federal services and to provide the public with greater ease of use, the government has structured its site according to broad client groups, known as portals or clusters, rather than along departmental lines. And, because Statistics Canada has information and statistics on many subjects, it has been called upon to participate in most of these clusters. As well it plays a leadership role in two clusters, *Business Statistics and Analysis* and *The Economy*".¹¹

Large and comprehensive "national" statistical web sites are available in other OECD countries¹² and in 2001 the OECD itself created a Statistics Portal on its new web site adopting a new presentation of statistics "by theme" rather than "by directorate". Because of the decentralised model adopted by the OECD for its statistical activities,¹³ the problems encountered in the development of the Organisation's

11 See "Statistics Canada and the Government On-Line Strategy", David Roy, *OECD Statistics Newsletter No.7* at <http://www.oecd.org/pdf/M00026000/M00026171.pdf>.

12 Examples include the United Kingdom where the National Statistics portal provides access to many other government websites and the United States of America where the FedStats portal provides a gateway to statistics from over a hundred Federal agencies.

13 At the OECD, various statistics are developed both by the Statistics Directorate and by Directorates responsible for analytical studies and policy analyses. The Statistics Directorate (STD) was created in 1992 with the following mandate: (a) to improve the supply of relevant and timely statistical information to analysts and policy makers inside and outside the Organisation; (b) to develop international statistical standards, systems and classifications in collaboration with other international statistical agencies; (c) to improve co-ordination between the statistical activities of the OECD and those of other agencies; and (d) to provide a mechanism for co-ordinating statistical activities within the Organisation.

From a substantive point of view, STD is responsible for macroeconomic statistics (national accounts, short-term economic indicators, international trade, etc.) and for some social (i.e. labour force) and business statistics. In addition, it plays a key role in promoting internal co-ordination and co-operation with other international organisations. The majority of other statistical activities are carried out in eight Directorates: Economics (ECO), Employment, Labour and Social Affairs (ELS); Education (EDU); Science, Technology and Industry (STI); Financial, Fiscal and Enterprise Affairs (DAF); Environment (ENV); Public Governance and Territorial Development (GOV); Development Co-Operation (DCD).

The Service for Information Technology and Network Services (ITN) and the Public Affairs and Communications Directorate (PAC) play an important role in supporting the development of OECD statistics. The former co-operates with STD and other Directorates to develop statistical databases and other IT infrastructures for conducting statistical activities, while the latter is responsible for the dissemination of all OECD products, including statistical data and publications.

The governing board of the OECD is the Council, which comprises the official representatives of Member countries and establishes the general policy and priorities of the Organisation, in close contact with the Secretary General. Directorates also support one or more Committees, who represent national governments and establish priorities for their own future work, evaluate reports prepared by the Secretariat, develop recommendations, etc.

Statistics Portal are very similar to those encountered by NSOs involved in the creation of national one-stop gateways to official statistics. In a sense, looking at the experience of the OECD, it is possible to identify typical issues for this kind of initiative, such as:

- the development of common classifications and definitions across different data providers;
- coherence among statistics produced by different producers;
- the establishment of common rules in disseminating statistics to the public.

The first issue is to develop a comprehensive list of “domains”, where common statistics compiled by various data providers can be presented. This task is quite difficult, because:

- the classification of statistical themes developed by the United Nations (UN) is not widely recognised and accepted;
- very heterogeneous approaches are applied to presenting thematic statistics on the web sites developed by NSOs and international organisations;
- it is not always obvious to users where to look for the statistical data they require especially if they are unfamiliar with a website and are “searching” for an item rather than “navigating” the site as regular users would do. A common problem is that the search engine on the website often fails to recognise certain key words, preventing users from finding what they need. Accessibility problems can also arise if a data provider uses an excessive number of levels through which a user must “drill down” in order to access the data. Thus, the classification should be as flat as possible but this may not be easy to do across different public agencies;
- domains can be defined following different criteria and several statistics are multipurpose, which means that the same statistical product could be classified to several domains;
- regular users can react negatively to radical changes in classifications previously adopted.

In the OECD experience, a good starting point has been the UN Classification of Statistical Themes, which has been adjusted to OECD use taking into account the variety of statistics provided by various Directorates. In particular, a three-level classification has been adopted, based on 23 first-level items, divided into 109 second-level items, which contain the various publications, reports and statistical data that comprise third-level items. This classification system enables the presentation of single statistical activities in the OECD integrated statistical program, providing users with a unique classification to obtain information about statistical processes and outputs (publications, databases, etc.).

The second, and much more difficult, issue is that different data providers can release publications and databases containing the same variables, but with different figures. For example, for the variable “employment” it is possible to derive figures from both administrative and statistical sources, and from the latter, estimates can be obtained:

- from surveys carried out on establishments or enterprises;
- from surveys carried out on households (i.e. labour force survey);
- from national accounts,

To deal with technical issues, Committees establish working parties or task forces. Therefore, several statistical bodies have been established over time.

To ensure the internal co-ordination of statistical activities, the Statistical Policy Group (SPG), representing all Directorates, was established in 1990. The SPG holds monthly meetings and is chaired by the Director of STD, who is the Chief Statistician of the Organisation.

using definitions that in several cases can differ from each other. In addition, if files (as normally occurs) are released at different moments in time, they reflect information at a particular time and are influenced by the normal revision process of statistical data. From a user perspective, the presence of inconsistencies across different databases (or publications) can be very confusing, even if the data are released with appropriate metadata. In fact, users often use search engines to find files containing data and as these files may contain only raw data, they may not have access to the metadata that can explain inconsistencies.

This is another problem that can be alleviated by portals since they provide integrated services and information across administrative boundaries so that users do not need to know which agency is responsible for providing the service or information. Thus, statistics portals can offer statistical information to users without them having to know the agency that produces them. One such portal is FedStats, which provides access to statistics produced by more than 100 United States Federal agencies. This was developed by the Interagency Council on Statistical Policy under the Chief Statistician of the United States to address barriers to accessing Federal data.¹⁴

Nevertheless, it is clear that, in the short run, the issue of consistency can be managed only through the dissemination of appropriate metadata that explain sources, methodologies and other statistical processes used to produce each data set, helping users to choose the best data for their specific purposes. In the long run, better integration between different sources can be based only on the adoption of harmonised classifications, variables, statistical methodologies, etc., or on the provision of expanded satellite accounts, usually developed by collaboration between different statistical units within an agency to solve inconsistencies. Take, for example, the labour accounts developed by the Netherlands Central Bureau of Statistics. Before revision, users encountered difficulties in linking national accounts figures with labour accounts data owing to differences in definition, population, methodology and time. Now the two sets of accounts are linked, consolidating two previously distinct processes and improving overall quality. Other countries such as Finland have followed the Dutch example.¹⁵ It is worth stressing that while ICT facilitates this type of work, a basic willingness within the organisation to embrace cultural change is also required.

The third issue, the establishment of common rules for disseminating statistics to the public, is even more difficult. On the one hand, it involves both institutional and political aspects, depending on the organisation of national statistical systems and the role of NSOs vis-à-vis the rest of the public sector. On the other hand, it is related to the policy of access to public information adopted by political authorities and to the management of an “information society”. In fact, the development of ICT techniques, the competitiveness between private and public sources and the launch of national e-government policy initiatives are key factors that have pushed statistical authorities to change their previous dissemination policies. These policies have altered radically over the last decade, with accelerated change towards a new vision in this area having taken place in the last two to three years.

From an institutional point of view, the power of a central authority in establishing common rules to disseminate statistics produced by different public institutions depends on the legal framework existing

14 The United States statistical system is decentralised, with individual agencies being responsible for statistics in their field of competence. Their activities are co-ordinated to a certain extent by the Statistical Policy Office in the Office of Management and Budget, Executive Office of the President, which is also responsible for ensuring that the budget proposals of individual agencies are consistent with government's overall priorities concerning the quality of Federal statistics. For more detail see “Report on Statistical Programs in the United States”, Suzann Evinger, *OECD Statistics Newsletter No.8* at: <http://www.oecd.org/pdf/M00027000/M00027530.pdf>.

15 For more detail see “Linking Social and Economic Statistics through the 1995 Revision of National accounts and Labour Accounts”, Statistics Netherlands, at: <http://www.cbs.nl/nl/publicaties/publicaties/macro-economie/nation-ale-rekeningen/occpaper93.pdf>.

in that country. Even in centralised systems, the co-ordination of data dissemination across public administrations is very difficult and succeeds more often through dialogue and co-operation than through the imposition of strict rules and bureaucratic processes. In this area, the initiative taken by the IMF to improve the quality of national statistics through its Dissemination Standards Bulletin Board (DSBB) represents an important step forward.

In particular, the establishment of the IMF Special Data Dissemination Standard (SDDS), which highlights both the “quality” of national statistical systems and standards of best practice in four areas - data coverage, periodicity and timeliness; public access to the data; integrity of the data; and data quality characteristics such as accuracy, adherence to international statistical guidelines, and consistency - for the dissemination of a selected set of statistics relevant to economic analysts and policy makers¹⁶ is changing the approach to the dissemination of statistics adopted in OECD countries. To meet the IMF quality standard, a country has to inform users about how it complies with best practice and compile a national page on the SDDS web site following a predefined format. This information consists of methodological statements (covering analytical framework, concepts definitions and classifications; scope of the data; accounting conventions; nature of basic data; compilation practices; and other aspects) and information that permits cross checks for reasonableness. The transparency of the statistical processes and definitions used by national agencies in the compilation of the statistics is the key to the whole process.

The IMF standards are increasingly used by national data providers to disseminate statistics other than those included in the SDDS list. In addition, because the IMF asked national authorities to nominate a single “national co-ordinator”, various data providers within a country had to reach agreement among themselves. Normally, this role has been attributed to NSOs, thereby strengthening the co-ordinating role of the statistical office, which, in general, is also in charge of the co-ordination of national statistical systems. The existence of an international standard has helped NSOs to push other public bodies to adopt similar rules, without opening a long and difficult negotiation process.

2.3. *Towards free dissemination of official statistics*

As already noted, the development of ICT techniques, the competitiveness between private and public sources and the launch of national e-government policy initiatives have encouraged statistical authorities to change their previous dissemination policies, in terms of timeliness, scope, classification etc. In the context of e-government in particular, with its policies of supporting the evolution of national societies, improving the efficiency of the public administration system, minimising “digital divide” among citizens, etc., statistics are fundamental in filling the information gaps between different parts of society and to improving the decision-making processes of public bodies, businesses and individuals. As a result, many statistical authorities have changed from a dissemination policy based on charging for access to statistical sources (publications, databases, CD-ROM etc.) towards the provision of free data.

From the perspective of information being a fundamental instrument to improve individual and collective wellbeing, the establishment of official statistics as a key “public good” element has been reinforced. This vision has provided a catalyst for statistical offices to substantially enlarge their distribution of free statistics on the Internet. Even in those countries where rigid and strict rules once were applied in disseminating statistics, almost all “basic” statistics¹⁷ are now disseminated free on the Internet. In some countries, this change has been presented as part of the more general national e-government policy

16 They cover various domains, such as national accounts, financial statistics, general government statistics, money and banking statistics, etc. In general, at least three different data providers are involved: the national statistical office, the national central bank and the Treasury ministry.

17 This generally means commonly published statistics for each subject matter.

and the loss of revenue for NSOs due to this passage has been fully compensated by governments, through an increase of public allocations to statistical offices.

In 1999, Statistics Denmark carried out a review of dissemination policies in seven European countries: Finland, Sweden, Norway, Denmark, The Netherlands, United Kingdom and France. The main conclusions of this comparative study are the following:

- in all surveyed countries all publications that can be downloaded from the web are given free of charge;
- in Sweden, Norway and the Netherlands the step has been taken to give first priority to dissemination of publications via the web, second priority to paper publications. This means that the web version is regarded as the “real” publication, the paper version is an extra service;
- the PDF or portable document format is generally considered to be the most suitable format for disseminating large and static publications containing complex tables and graphs;
- there is a tendency to substantially enlarge the availability of databases free of charge. Databases can be password protected even if they are free of charge;
- in Sweden (where statistics on users are available), after the move towards free of charge databases, the number of users increased from 500 (December 1999) to 10,000 (December 2000) and yearly retrievals increased from 72,000 to 160,000. Preliminary figures show similar tendencies in Denmark (50,000 retrievals in the first month following the launch of the free access database);
- a remarkable change in the composition of users occurred in Sweden after the decision to adopt free dissemination. In particular, ministries, local governments, etc. were the main users during the period of paid service, while more than 50% of all users are now private enterprises;
- metadata are very important, especially when users retrieve data themselves.

Several of these conclusions are also true for other OECD countries. The experiences of the US, Australia, New Zealand and Canada show similar results, and strategic plans to adopt new dissemination policies centred around the Internet are in progress. Finally, it is noticeable that in several countries, in the framework of e-government policies, NSOs have been given responsibility for horizontal projects to improve the electronic dissemination of data produced by other public bodies, following the same standards and policies already applied by statistical authorities.¹⁸

From the standpoint of international organisations, the Statistics Directorate of the OECD (STD) in 2001 reviewed Internet dissemination policies adopted by members of the Administrative Committee on Co-ordination (ACC).¹⁹ The review showed that in 2000 all ACC members disseminated at least some statistics via the Internet, though there was considerable variation in both the amount of data disseminated, and the amount of data disseminated free of charge. In particular:

18 Apart from the Statistics Canada example already mentioned, the Australian Bureau of Statistics has initiated a project called the National Statistical Service which asks other government agencies for a shared commitment to improving the availability of statistical information including data that can be generated as a by-product of administrative activities. Also, the United Kingdom Office of National Statistics has responsibility for making the web integral to National Statistics, which provides a centralised statistical knowledge base across the UK public administration.

19 The ACC was a UN body where all international organisations active in statistics were represented. The STD report commented on the Internet statistical dissemination practices of 23 international agencies.

- the International Labour Organisation (ILO), UNESCO, World Trade Organisation, UN Economic Commission for Africa (UN/ECA), UN Economic and Social Commission for Asia and the Pacific (ESCAP), UN Industrial Development Organisation (UNIDO) all applied a general policy of free distribution of statistics on Internet;
- Eurostat, Economic Commission for Latin America and the Caribbean (ECLAC) and the FAO provided extensive information free via limited selection/extraction facilities;
- UNICEF and the UN Statistical Division provided an extensive range of information free via readily accessible static pages, while others provided extensive information in the form of PDF files;
- in a number of instances the Internet site also served as a portal to extensive subscription on-line databases (Eurostat, World Bank, ILO, FAO, UN Statistical Division, World Tourism Organisation, UNIDO);
- the majority of methodological information (statistical guidelines and recommendations, publications outlining current statistical methodologies used by member national agencies, reference documents, etc.) were normally disseminated on Internet free as a matter of policy. The trend is to make more such information freely available on this medium in the future due to its high public good component, particularly with regard to the promotion of statistical transparency, best practice and the use of comparable statistical methodologies.

Concerning the OECD itself, it was considered to be applying a very restrictive policy in terms of free distribution of statistics owing to the fact that revenue from publications was an important factor in framing the Organisation's budget. This policy was quite different to those applied in many NSOs, where the character of official statistics as a key "public good" element was being recognised more and more.

Following consideration within OECD of the Organisation's dissemination policy, it was agreed that one objective with regard to dissemination of statistics should be to contribute to the development of a culture of "informed decision making" at national and international levels, both in government and non-government bodies. To meet these objectives, the statistics dissemination policy must - in accordance with OECD general publishing policy and with the financial and budgetary regulations and requirements of the OECD Council - ensure that

- the general user community has free access to basic statistical information and all metadata collected and/or originally produced by the Organisation;
- free access to all statistical products is given to all national governmental bodies, as well as, subject to reciprocal arrangements, to international organisations.

With the launch of the new web site a new, less restrictive policy was agreed so that for each page created on the site, a sample subset of available data now must be provided free to the general user, subject to a limit of a maximum of 10% of overall available data. These data are provided following standards fixed by STD, and utilising software already available at the OECD.

Even in those countries that disseminate statistics free on the Internet, books and other more complex statistical products are normally for sale. In some countries a print on demand mechanism, for books and CD ROMs has been put in place. Furthermore, differentiated prices are applied to different groups of users (public administrations, universities and schools, businesses, etc.). Countries which have commenced dissemination of data for free on the Internet have also had a simultaneous increase in demand for free data which has spilled over into increased demand for other products, though the net impact of these recent trends on total revenues is at yet unclear.

In summary, there is a clear tendency towards the major use of the Internet to disseminate statistics, and towards a policy of free data dissemination through this medium both at the national and international levels. This tendency is encouraging all NSOs and international organisations to carefully re-think their existing policies, to apply the most advanced techniques to improve the accessibility of statistics (and related metadata), and to investigate the efficiency of the process.

3. Statistical languages and models for the exchange of information

3.1 *Role of metadata in statistics*

Significant enhancements in the power of computers over the past few decades has led to enormous efficiency gains in the collection, compilation and dissemination of statistical information. Processes that took days a few years ago may now take seconds. Calculations that were once almost impossible are now easily done. Access to and transmission of data and information is much easier. The Internet has enabled statisticians to instantly distribute data at low cost over long distances to large numbers of users.

Unfortunately, these advances have also created opportunities for the provision of vast amounts of confusing information, as discussed in Part 2.2. It is important that in extending e-government services, governments avoid these pitfalls, at the same time taking advantage of the opportunities that these new developments offer. Methodological information must be provided in a way that enables a user to evaluate the quality of that data in terms of criteria such as coverage, accuracy and comparability. It should enable a user to focus on the definitions, sources and methods of compilation etc. of the indicators in question so that the relevance of a statistical value in relation to a specific need can be better understood.

It is interesting to note that in countries such as Australia and Korea, minimum standards for metadata to facilitate the discovery of e-government resources and services have been developed. Such standards help website managers in deciding on the content and design of their site so that its structure is more client-oriented. A common structure with mandatory elements, content rules, consistent encoding schemes etc. across portals and sites enables users to find more easily the services they want. Of course, this is an area where statistical institutions have a major responsibility to their clients.

Statisticians are – or should be – particularly adept in providing methodological information that is of use to a user. The primary role of the statistician, in relation to statistical methodological information, entails its collection, verification and dissemination. He/she also is experienced in giving it structure and providing a clear path that enables users to dig as deeply as necessary without being lost in enormous amounts of confusing or irrelevant text. In addition to helping others make use of statistical methodological information, statisticians also use it themselves in evaluations and assessments of data quality and comparability. Their experiences have made them very aware of the limitations of even the most detailed methodological information that supports a particular data series, whether they be limitations that stem from problems of accessibility, differences in semantics (the same term does not necessarily have the same meaning across different agencies) or the fact that national agencies frequently describe different aspects of the statistical production cycle.

Nevertheless, despite these possible limitations, with the development of e-government portals and fully integrated e-government services, exchange of information between government agencies is not an option - it is a necessity. Whilst there are concerns to be addressed regarding the security and privacy of the individual person or business making use of such services, in many cases the individual user gives permission for personal data to be exchanged between the various government agencies that may have

some specific responsibility in the overall programme or service. This may be an attractive option to the user when the administrative burden is reduced. However, on a technical level, such exchange is not easy because various agencies more frequently than not use different databases and IT systems.

The establishment of a metadata standard is generally seen as one of the best tools in overcoming these deficiencies. In this regard metadata does not mean methodological information but rather agreement on a structured set of statistical metadata items on the website. A metadata standard helps website managers to decide on the content and design of their site so that its structure is more client-oriented. A common structure across portals and sites enables users to find more easily the services they want. Such a structure also enables portals that provide "one-stop shop" access to services requiring multi-agency involvement but which the client sees as a single service.

Moreover, a common structure facilitates the exchange of data between government agencies. Recent decades have seen the development of statistical languages and standards that enable closer integration between agencies and quicker and more efficient exchange of information between them. However, the development of such metadata standards and, more importantly, their successful implementation requires a close alliance between IT specialists, statisticians and users. There are numerous examples of "failed" projects where such alliances have not been developed and utilised.

3.2 *New standards and statistical languages for exchange of data and metadata*

The advent of inexpensive electronic communications in the last quarter of the twentieth century led to the development of new technologies for electronic exchange of statistical and non-statistical information. For example, the 1990s saw the development of the Generalised Statistical Message protocol (GESMES), which facilitates the construction and deconstruction of messages containing statistical data and metadata, and also the Extensible Mark-up Language (XML), which was a response to the need for a language that addressed the content of information embedded in text as the amount of information on the World Wide Web expanded.

An improved message structure called GESMES/CB was introduced by the Bank for International Settlements, Eurostat, and the European Central Bank (and adopted by the IMF) toward the end of the decade. By the turn of the millennium, electronic exchange of statistical data had become standard business practice among these central agencies and their member countries. However, although national sources may provide very similar data to several international organisations, there are also a number of differences that make each data request unique. The details and processes of data collection of different international organisations vary considerably in response to varying needs. The nature of the details and processes involved depends on a number of criteria:

- nature of data to be collected, e.g. whether or not they are part of a regular production process;
- frequency of the collection, which varies from daily to annual or even occasional;
- regularity of the content over time;
- homogeneity of content regarding the internal structure of the NSO (e.g. subject matters);
- quantity and level of data (e.g., disaggregated versus aggregated data);
- method of collection;
- security and confidentiality requirements
- technology and infrastructure.

In addition to the diversity generated by those criteria, processes for data collection have evolved over the years as the result of bilateral arrangements between individual organisations and national agencies or other international organisations. Also, policies regarding arrangements vary from accepting any format and medium suggested by the data provider to legally imposing a fixed format and fixed medium. In particular, some regional and/or transnational agencies are more likely to prescribe fixed formats and medium. Eurostat, for example, recommends the use of a precise format, in general one of the implementations of GESMES. Eurostat also recommends the use of its data collection infrastructure, in particular the Stadium system which is a centralised service for collecting and administering statistical data files submitted to Eurostat. Stadium also redistributes incoming data files to the appropriate Eurostat areas. However, not all EU countries follow Eurostat's recommendations. The ECB, at the other extreme, requires that all data reporting uses GESMES/CB.

Despite many efforts to develop standards and to increase co-operation, the international statistical community currently uses an enormous number of data streams, formats and technologies to exchange statistical data and metadata. Media include paper, diskettes, CD-ROMs, mainframe magnetic tapes, email, other file transfer methods, web databases and other on-line databases. Formats vary from text with separated or fixed fields (e.g. CSV), proprietary formats such as MS-Excel or Fame databases, on-line questionnaires and of course the statistical standard GESMES in its various versions. The structure of the data received varies from tables, sets of times series and predefined questionnaires.

Nevertheless, despite these limitations, there have been significant developments in standardisation of data exchange in recent years. Following the pattern for data, the newly developed sets of metadata are also being exchanged between and among national states, regional and international organisations, and the general public. The need for standardisation of metadata exchange is a logical outcome of the increasing need to exchange metadata.

Although the development of standards for metadata exchange is still in its infancy, there are several examples of advances in this area in recent years. As already mentioned, the IMF SDDS was introduced in 1997. In 1999, Eurostat introduced Euro indicators, a collection of data and metadata covering the euro-zone and EU-15 in the wake of the European Monetary Union.²⁰ In early 2001, the Euro indicators were pulled together into a single web site, where metadata are shown in SDDS format. Many OECD Members have also developed their own web sites containing a mix of data and SDDS metadata. In addition, the OECD has been working with other international institutions towards the development of more efficient processes in this regard by exploring common e-standards and ongoing standardisation activities. This is known as the Statistical Data and Metadata Exchange (SDMX) initiative.

3.3 *SDMX initiative*

At the concluding session of the Workshop on Statistical Data and Metadata Exchange, held at the IMF in September 2001 and attended by more than 100 participants from all over the world, it was recommended that the sponsoring institutions lead an international initiative to create the standards aspired to in the following statement that had been distributed to participants prior to the workshop:

“The BIS, ECB, Eurostat, IMF, OECD, and UN have joined together to develop new common models for business practices in the field of statistical information that would allow more efficient processes for exchange and sharing of data and metadata within the current scope of our collective activities. These models will support the creation of common e-standards and allow us to avoid duplication of

20 See <http://europa.eu.int/comm/eurostat/Public/datashop/print-catalogue/EN?catalogue=euroindicators>.

effort in our own work and possibly for the work of others in the field of statistical information.”

Following this, the six sponsoring institutions established a taskforce to address the issue of Statistical Data and Metadata Exchange (SDMX).²¹

The SDMX initiative endeavours to create common standards that will suit the needs, not only of the sponsors, but also their member states and their data user communities. The work of this group entails a common set of guidelines and recommendations that will enable participating institutions to take maximum advantage of exchange protocols such as GESMES/CB and e-standards such as XML. Thus, there is a need to create an open and transparent process for participation of member states and data/metadata consumers in the development of the standards. The idea of an open process centres on a few key ideas. These are:

- all parties interested in engaging in the effort to create a given standard and willing to provide their own time and effort may participate;
- the cost of participation should be born by the participants;
- the cost of participation should be minimised to the extent that it is not a significant barrier to willing participants;
- the intellectual property developed by the process should be freely available for public use at no cost;
- the process should be governed by a formal democratic process; and
- the deliberations taking place within the process should be archived and publicly visible.

The results produced by an open process should be based on three technical principles. These are:

- the structure should be captured in a standard way so that it can be used by any tool or technology and not be dependent on a specific vendor’s product;
- the structure should be described in a language that is extensible, allowing for additions as new information is created; and
- the language used to describe the structure should be independent of formatting and presentation features, thus allowing these features to be determined by each user.

Both the GESMES and XML vehicles fully conform to these three principles.

SDMX is attempting to establish standards for the direct exchange of files between parties, for the placement of data and metadata on a web site that can then be accessed by users, for describing data and the metadata associated with it, for describing catalogued metadata such as that in the OECD sources and methods publications and the IMF DSBB, for presenting time series data and presenting tabular data, etc. By using the ideas referred to above as guidelines for the development of standards for data and metadata exchange, it is expected that barriers to the sharing of the intellectual property developed by SDMX will be minimised. In addition, these ideas are intended to encourage the widest possible adoption and to encourage the marketplace to develop products that support usage of the standards created.

The heads of statistics of the six sponsors met in June 2002 and agreed on four projects that would advance the SDMX initiative. These four projects were:

²¹ See <http://www.sdmx.org> for the common statement by these organisations.

- a practical case study on emerging e-standards for data exchange;
- maintaining and furthering existing standards for data exchange of time series;
- creation of a common vocabulary for statistical metadata;
- development of a framework for metadata repositories.

The objective of the practical case study is to investigate the ways in which emerging web technologies can be used to standardise and simplify the collection, compilation and dissemination of statistical information and to identify the core set of e-standards that are required to achieve the full potential of the emerging web technologies such as XML. It is intended that the project will demonstrate exactly where these new techniques offer improvements over existing methods of conducting business both now and in the medium-term.

The goal of the second project is to maintain the use of GESMES/CB by existing users and to encourage its use among potential users, particularly where this is to the benefit of SDMX organisations. It will develop more formal open structures and procedures for the furthering of GESMES/CB, based on best practices for standard-setting at the international level.

The aim of the third project is to produce a common vocabulary of key metadata items that describe statistical concepts and methodologies used by statistical organisations. It is intended that the vocabulary will be used in existing dissemination frameworks such as the IMF SDDS. Metadata that is at least consistent with regard to content between countries is essential if metadata provided by different organisations is to be used for comparing national statistical practices. The vocabulary uses definitions, etc., from existing international statistical guidelines and recommendations. It will be a subset of a more comprehensive glossary, jointly developed by OECD and Eurostat that contains a large number of definitions on statistical terminology and concepts as well as definitions of variables commonly used in economic and social statistics.²²

The objective of the metadata repositories project is to facilitate and widen the exchange of metadata among international organisations and across countries by standardising the way in which repositories of metadata are organised. The ultimate objective is to standardise the search vocabulary of national and international agencies and even private sector institutions so that a single query can be sent to multiple sites at once.²³

4. Statistics and confidentiality

4.1 *The key issue of confidentiality for statistics*

In response to the pressures of declining budgetary allocations and increasing demands from business to reduce the administrative burden that statistical compliance places on them, NSOs are attempting to decrease enterprise reporting burden. One of the ways in which this is being done is through

²² The more comprehensive OECD Glossary of Statistical Terms is located at: <http://cs3-hq.oecd.org/scripts/stats/glossary/index.htm>. This glossary currently contains around 5,500 definitions.

²³ For more information on these projects, see “Statistical Data and Metadata Exchange (SDMX): Moving Forward”, *OECD Statistics Newsletter No.11* at: <http://www.oecd.org/pdf/M00034000/M00034959.pdf>.

greater use of administrative data in lieu of direct collection.²⁴ The UN Handbook, *The Operation and Organisation of a Statistical Agency*, encourages this approach from an efficiency perspective. However, the use of administrative data has far reaching implications for NSOs as to how they deal with confidentiality concerns.

Although statistical data are often collected as a by-product of the administrative records of other public bodies such as tax collectors, customs authorities, etc., NSOs do not provide policy-makers with statistical records at the level of the individual entity other than enabling them to use anonymated data at an aggregated level.²⁵ Given the necessity to protect the independence and integrity of national statistical institutes, there is no reason why statisticians should relax their traditional emphasis on statistical confidentiality even though technological developments in software and communications have resulted in major changes regarding the collection and compilation of data, exchange of information, and dissemination policies. Indeed, technological advances have afforded not only the opportunity for enhanced security procedures but in some cases, such as data warehousing, they require them. In addition, the World Wide Web has also enabled statistical institutes to publicise and reinforce their commitment to statistical confidentiality before a wider audience than was possible previously.

Statistics are provided to enable users to analyse data and/or make informed decisions. It follows, therefore, that the activities of a national statistical institute should be primarily focused on the needs of the user. On the other hand, indiscriminate disclosure of all micro data to users by a statistical institute would have an enormous negative effect on the willingness of respondents to co-operate in any of that institute's statistical programmes. So there is a trade-off between protecting the privacy of respondents from whom data are collected and providing users with enough data for their own particular needs.

This is not an easy balance to achieve. The experience of statisticians with regard to user demand is often that they provide the data too late, with insufficient coverage, outdated classifications and poor standards of comparability. The experience with regard to respondents' perceptions is that the political independence of national statistical institutes is often questioned and disbelieved. In addition, respondents often complain that undue burden is placed on them to provide data that might be of no great importance to anybody other than a few academics. These perceptions may be unfair but they exist.

Despite complaints from both sides, however, the bottom line is that statistics must be provided in a level of detail that enables users to use them but must also be collected and disseminated having regard to the privacy of the initial respondents. If the confidentiality aspect is ignored, the level of response may be so poor that the data provided as an output to users is so unreliable as to be useless. As the United States Office of Management and Budget has stated, "Public confidence and willingness to co-operate in statistical programs substantially affects both the accuracy and completeness of statistical information and the efficiency of statistical programs."²⁶ It is, therefore, in a national statistical office's self-interest to protect micro data that respondents provide. This is also one of the United Nations Statistical Commission's Fundamental Principles of Official Statistics.²⁷

24 At the OECD Short-Term Economic Statistics Expert Group (STESEG) meeting in June 2002, several NSOs reported that they were making increased use of administrative data. The use of administrative data will also be considered by the Group in its future activities.

25 In some countries, individual statistical records can be transmitted to other public administrative bodies but they can only be used for statistical purposes.

26 *United States Federal Register*, Volume 62, No.124, 1997.

27 For more details on these principles, see <http://unstats.un.org/unsd/goodprac/bpabout.asp>.

Acquiring a person's agreement to voluntarily provide personal data for research and statistical purposes is better than forcing them to do so under fear of penalty for failing to do so (although statutory authority is usually considered necessary to support this). Statistics offices have spent, therefore, large amounts of resources seeking to gain informed consent, including the publication of confidentiality policies through the Internet. Although policy details will differ from country to country, there are a number of common principles that apply. The main principles are:

- data should be used only for statistical compilation and analysis;
- information should not be disseminated in a form that can be related to an identifiable person or business entity, save for specified exceptional circumstances;
- there must be steps to avoid statistical disclosure (ensure that no individual entity can be recognised from disseminated data);
- there should be a specified punishment in the event of the law being broken.

Statisticians use different variations of legal, administrative, methodological and technical measures to collect and disseminate data in a manner that fully addresses respondents' concerns. Some of these methods are outlined below both in a general context and in the context of e-government.

4.2 *Legal and administrative measures for confidentiality*

In OECD countries, the legislation under which national statistical institutes operate contains specific reference to protecting the confidentiality of individual responses. Of course, cultural differences mean that protection measures with regard to disseminated data may vary from country to country. For instance, some countries may permit greater flexibility than others in exceptional circumstances for data that is not normally disseminated. However, all recognise the importance of confidentiality, in line with the UN Fundamental Principles.²⁸

In addition, most countries have also introduced legislation aimed specifically at controlling misuse of electronic data (e.g., *Data Protection Act* in Ireland, *Privacy Act* in Canada). These laws, which apply to any entity maintaining personal or identifiable data, reinforce statistical legislation in the context of statistical confidentiality. Depending on the specific legislation, users of data provided by the national statistical authority could also be obliged to protect the privacy of individual entities that furnished the original micro data. The strength of such legislation lies in the fact that any breaches of confidentiality are punishable under the law.

Of course, many citizens are suspicious of government and they may have little faith in legislation as a protective barrier against direct or indirect statistical disclosure by a national authority. So, although the law may offer adequate formal protection to citizens and business, additional measures may be necessary to assure respondents that the national statistical authority is fully committed to protecting their confidentiality as a matter of corporate policy and not just because it is legally obliged to do so. In this regard, it could be beneficial to develop administrative measures such as a voluntary code of practice to underpin legal measures and to emphasise that it will not just do the "bare minimum" as required by law. In other words, a code of practice can reinforce assurances afforded the respondent under law. Confidentiality pledges assure respondents that information provided will not be used against them in any government action. As already stated, the Internet enables institutes to advertise their commitment to confidentiality, whether through international organisation metadata dissemination formats such as the SDDS or through

²⁸ Confidentiality is also a requirement of the SDDS. See <http://dsbb.imf.org/sddsindex.htm> for detailed information from subscribing countries regarding the subjects included on the DSBB.

their own web sites. One example is the National Statistics Code of Practice in the United Kingdom which was introduced in October 2002 following widespread public consultation. Interestingly, the Code applies not just to the Office of National Statistics but to all statistics classed as “national statistics” and it is to be applied to all public sector statistical work.²⁹

The increased levels of co-operation between statistical authorities and other government agencies with regard to administrative data may also arouse public suspicion. The fact that the national statistical authority can obtain data³⁰ from other public bodies does not imply that micro data flows in the opposite direction (although exchange of anonymised statistical datasets are generally permissible). However, there is sometimes a misplaced perception that two-way exchange is the norm, with the result that some respondents are reluctant to provide data that they think could be used by other government authorities such as taxation or social security offices. Again, statistical offices are aware of the need to emphasise that such is not the case and that exchange of data only takes place within a context that fully protects the confidentiality of individual entities.

Another problem with use of administrative data for statistical purposes is that the statistical element is only secondary to the primary reason for collecting the information. If data are not collected for statistical reasons only, there is a greater possibility that protective measures for non-disclosure will be circumvented. For instance, in themselves, anonymous statistical data published by an NSO and confidential personal data that is maintained on another public body's database may not present any danger of disclosure. However, comparing NSO published data and non-confidential elements of the other register may enable a determined person to discover confidential details. With e-services making it more attractive from an efficiency and effectiveness perspective to use administrative data for statistical dissemination and analysis, statistical authorities must have regard to this potential problem. This is known as secondary disclosure and some methods used to minimise the danger of this are outlined in Parts 4.3 and 4.4.

In some countries, special procedures apply to researchers. Although data may not be released to the general public, researchers may be given special privileges to enable them to do their work. Respondents should be made aware that these arrangements apply and the statistics authority should also do its utmost to prevent abuse of these privileges. An interesting approach to disclosing data exists in Sweden. In that country, it is a crime to try to disclose personal data. This makes it much easier for the statistics office to release anonymous micro data files since in the case of breach of confidentiality, responsibility rests with the entity that divulged the data instead of the office that released the original file. In Norway, datasets are released to researchers on a contract basis, the identifier and geographical fields first having been removed. The statistics authorities in both Sweden and Norway have developed several approaches towards combating possible abuse of privilege by researchers.³¹

29 Ireland is another example of a country where the national statistical institute has a well-advertised code of practice concerning confidentiality although the code does not apply to other public sector statistics. For more detail see <http://www.cso.ie/misc/confident.html>.

30 This process is very advanced in Sweden, with administrative sources accounting “for more than 95% of the statistical micro data used for official statistics production” according to Working Paper No.25 “Statistical Microdata - Confidentiality Protection Versus Freedom of Information”, Statistics Sweden, Joint ECE/Eurostat Work Session on Statistical Confidentiality, March 2001.

31 See Joint ECE/Eurostat Work Session on Statistical Confidentiality, Working Paper No. 25, “Statistical micro data - Confidentiality Protection versus Freedom of Information”, Statistics Sweden, March 2001 and Joint ECE/Eurostat Work Session on Statistical Confidentiality, Working Paper No. 41, “Confidentiality Problems Related to Survey Data in Norway and Some Possible Approaches”, Statistics Norway, March 2001.

It should also be noted that researchers themselves and other users of data may have concerns with regard to privacy and confidentiality. To allay these concerns, agencies should be transparent about the need for a user to reveal personal information. For instance, the Statistics Canada web site contains information that explains the circumstances in which personal information is required as well as guaranteeing that it will not be provided to outside bodies. It also provides the client with the opportunity to inform Statistics Canada if he/she does not wish to be contacted again for promotional and/or market research purposes. Furthermore, it refers to software programmes that monitor traffic to identify unauthorised attempts to access or change information - and states that it will not attempt to identify individuals visiting its site unless an attempt to damage the site has been detected. Privacy notices are found on all order forms, invoices and on the web site. Likewise, the Australian Bureau of Statistics contains a Website Privacy Statement that informs users of the personal data that is recorded and how it is maintained, so that the user can use other options if he or she does not have confidence in the electronic system.³²

4.3 *Methodological and technical measures for confidentiality*

Of course, these days, statisticians must - along with any other entity that uses computers - ensure that access to computer files is limited by password and other physical security features. Any identity-disclosing information must be concealed through use of special impersonal codes. Two methods that have been used in Statistics Sweden are (i) the transformation of commonly known identifiers into non-identifiable variables and (ii) the development of a statistical firewall that filters queries from users as well as the statistical outputs resulting from these queries, thereby monitoring the flow of internal databases containing sensitive information to external users.

Before statistics are released they must be carefully analysed to avoid statistical disclosure. This means that care must be taken to avoid publishing results in such detail that users can make individual identifications. For instance, if there were only two pig farms in a particular area for which the total number of pigs was published, it would be possible for each to know what the other had. This is particularly possible where the scope of the population is small or where the distribution of some variables is skewed. The aim of disclosure control is to hinder re-identification of individual respondents from the published data and to hinder the derivation of information from the data, which is not supposed to be available.

In the previous paragraph, the word "hinder" is used rather than "prevent". This is because statistical institutes can only seek to minimise the possibility of personal information being disclosed.³³ There is no such thing as zero-risk of disclosure - the collection of data entails some risk no matter how small. A particular project dedicated to minimising this risk was the Statistical Disclosure Control (SDC) project in which the Netherlands Central Bureau of Statistics, the United Kingdom Office of National Statistics and Italy's ISTAT were among the institutions that participated. The work of that project is now

32 Mention should also be made of another concern of users that requires technical protective measures - providing credit card details. The move towards free dissemination of large amounts of statistical data notwithstanding, many statistical agencies now sell data over the Internet that is specific to an individual user's requirements and they make appropriate arrangements to ensure that these sensitive details are protected with encryption and other technical safeguards. In addition, it is advantageous if the protection policy and measures are explained transparently.

33 One of the issues relating to micro data identified as being of some concern to Members in advance of the UNECE Seminar on Statistical Confidentiality in February 2003 was whether automatic confidentiality protection was possible at all.

being continued under the Computational Aspects of Statistical Confidentiality (CASC) project, which seeks to build on the results of SDC.³⁴

The main output of the SDC project was the development of the ARGUS software package for both micro data (μ -ARGUS) and tabular data (τ -ARGUS). This package facilitates the modification of a detailed micro data set so that another set is obtained with acceptable disclosure risk and with minimum information loss. Given the difficulties associated with defining the risk of disclosure and quantifying information loss, it is clear that the availability of specialised software such as ARGUS has facilitated major efficiency gains by replacing the tedium of manual calculation and application of the rules with an automated process. In addition, it also ensures a degree of accuracy that is not possible with manual calculation.

It should be noted that in order to hinder re-identification, those variables that can assist in re-identification must be specified. This is a matter of judgement and so some minor scope for disclosure remains. However, once a variable has been specified as identifiable, it is an automated procedure to deal with it through ARGUS using options such as:

- global re-coding where several categories of variable are collapsed into one;
- the application of local suppression where one or more values in an unsafe combination are suppressed;
- use of a dominance or concentration rule³⁵ to suppress sensitive cells in tables, *i.e.* cells that cannot be published as they would reveal sensitive information on individual entities.

4.4 Innovation and regard for confidentiality

The *UN/ECE Meeting on the Management of Statistical Information Technology* in February 2001 agreed that "the desire to speed up innovation can lead to proceeding prematurely while overlooking critical aspects such as the careful design of the systems, continuity of existing practices etc."³⁶ This is a particular danger and it is of course important that new innovations are accompanied by new confidentiality arrangements where necessary.

One recent innovation has been the development of data warehouses. These store large amounts of data in an integrated database. Their management is not the responsibility of a particular data-producing division or section and there is a great deal of co-ordination and exchange of information. The opportunity to combine data from different sources has made it necessary to develop new confidentiality protection measures. However, database management systems often do not include tools for protecting confidentiality and this is a serious weakness that must be addressed. It is important to ensure that an agency's reputation is not damaged by intruders achieving disclosure or by inadvertent disclosure through legally available information.

34 For more detail on these projects and on ARGUS, see <http://neon.vb.cbs.nl/casc/default.htm>.

35 "This rule states that a cell is unsafe for publication if the n major contributors to that cell are responsible for at least p percent of the total cell value. The idea behind this rule is that in unsafe cells the major contributors can determine with great precision the contribution of their concurrents. In ARGUS the default value for n is 3 and the default value for p is 70 %, but these values can easily be changed if the user of the package prefers other values." Joint ECE/Eurostat Work Session on Statistical Confidentiality, Working Paper No. 2, "Statistical Disclosure Control of the Statistics Netherlands Employment and Earnings Data", Statistics Netherlands, March 1999.

36 See Report of the Meeting, Page 3 at <http://www.unece.org/stats/documents/ces/ac.71/2001/2.e.pdf>.

Sometimes, this can be difficult. For instance, some problems can only be resolved through expansive and time-consuming legislative changes. Also, current techniques may not be powerful enough to handle all the problems in large data warehouses (although the *Report of the March 2001 Joint ECE/Eurostat Work Session on Statistical Data Confidentiality* found that ARGUS should be powerful enough to deal with tables of any size or complexity). Only technical experts can provide the technical solutions but it is responsibility of statisticians to be aware of the problems so that the solutions can be developed.

Another problem that is being given increasing consideration is that of secondary confidentiality where confidential cells may still be calculated or derived from different tables containing different data on the same aggregation or geographical level. This type of data manipulation has become much easier with the availability of computers, statistics portals and other electronic advances and it is a problem. Statistics Sweden's development of the statistical firewall to filter queries (referred to in paragraph 74) is one such way of minimising risk. However, it is worth repeating that risk can only be minimised not eliminated. For instance, the Australian Bureau of Statistics specifically states that it can only protect its own data and admits that even where primary confidentiality exists, "information that may allow users to estimate the suppressed data may be available from other organisations".³⁷ Both the Swedish and Australian examples may provide options for resolving the problem encountered by some NSOs whereby legislation makes them responsible for secondary disclosure made possible by legally available primary data.³⁸

A combination of measures is probably the best way to maximise confidentiality, as argued by Statistics Sweden at the *Joint ECE/Eurostat Work Session on Statistical Data Confidentiality*. For instance, legal measures may not be enough as sometimes it can be too easy or tempting to break the law if the fear of punishment is outweighed by another factor. Although methodological tools such as ARGUS and firewalls are being developed all the time, there is always the danger that they can be beaten by a particularly determined "hacker". Even if there is no unauthorised access, methodological tools cannot provide complete protection as they can only work within the parameters set by statisticians' judgement. Moreover, the general public may not understand the complexity of a statistical office's methodological protection measures so that it does little to boost its confidence in the institute's commitment to confidentiality. So a mix of measures obviously works better than reliance on a single type of measure.

Technical tools that support the implementation of legislative, administrative and methodological tools can also be used to speed up the dissemination process as well as enhancing confidentiality measures. Where the process of controlling statistical disclosure was once slow and tedious, technological advances mean that the process now can be completed much more quickly and with less likelihood of error. The impact of new technological developments in software, communications and computing on statistical disclosure control was a major topic at the March 2001 Work Session referred to above. The Report of the Work Session stated that the increasing tendency to provide customised on-line access to databases was a "real SDC challenge". It found that data should only be released if it passed disclosure limitation rules with a query filter rejecting queries that did not pass such rules and a results filter checking output tables before allowing the client to access it. Both filters should take account of previous queries by the user to see if they are accumulating knowledge through successive queries so that they can re-identify an individual or business entity. Another effective method that the report mentioned was record linkage whereby disclosure

37 <http://www.abs.gov.au/ausstats/abs@.nsf/0/A2A945757C832F9ACA256A5B001BD7ACOpen&Highlight=0,confidentiality>.

38 It is envisaged that Statistics Sweden will lead a one-day seminar at the Conference of European Statisticians in June 2003 on "challenges for traditional approaches to confidentiality protection over the next ten years".

risk is empirically assessed using the degree of linkage between records in an SDC dataset and the original dataset.³⁹

³⁹ The report of the Work Session is available at <http://www.unece.org/stats/documents/ces/2001/31.e.pdf>.

5. Conclusions

E-government can be defined as the use of information and communication technologies (ICT), and particularly the Internet, as a tool to achieve better government. Therefore, the ultimate objective being better government, it stands that e-government is more about government than about “e”. As stated in the introduction, it should be about taking advantage of the opportunities presented by ICT developments to deliver higher quality public services to a wider community, which can then make objective use of information and statistics to evaluate the effectiveness of government, ultimately enhancing the concept of government being accountable to citizens.

However, in many instances, e-government policies have proved harder to implement in practice than in theory. Predicted efficiency gains have not been realised, costs have exceeded expectations, public servants have been resistant to the cultural and administrative changes required, and citizens and business have proved reluctant or unable to make use of ICT in their dealings with government (indeed the expansion in e-commerce in general has been weaker than expected). While e-government, through the enhanced capacity to collect and share information, can change the traditional ways in which government operates, too many administrations are clinging to the old models in the face of the new developments. If e-government initiatives are to be successful, organisations must make strategic, administrative and cultural reforms.

In this paper we have seen how various statistical organisations have used ICT advances to make efficiency gains in the area of public service delivery for which they are responsible, *i.e.* the collection, compilation and dissemination of statistics. Nevertheless, it is important to acknowledge that the successes achieved over the past decade in both national and international statistics organisations have been as much due to issues such as leadership, commitment, planning, involvement of many different actors, etc., as to the use of ICT. As mentioned previously, there have been examples of statistics projects that failed owing to poor planning or failure to involve all the relevant parties. On the other hand, the numerous examples of successful projects show what can be achieved as long as there is committed, informed leadership and a clear sense of the ultimate objective.

Areas in which national statistical authorities have used advances in ICT to improve the service offered to clients include:

- using the Internet to disseminate data more cost-effectively to the extent that they can offer tailored products updated in real time;
- using the expanding range of statistical tools to elaborate large masses of elementary data and to make ready-to-use products available to clients;
- enhancing integration with other public bodies to use administrative sources of data to produce efficiency gains that ultimately benefit users as well as respondents whose response burden is reduced;
- responding to the requirements of international bodies and complying with international standards regarding both data and metadata;
- disseminating more data free of charge in recognition that official statistics can often be a key public good.

All of these changes produced, and are producing, relevant changes in which NSOs operate and interact with data and metadata providers and final users.

In addition, NSOs have been actively involved in national e-government projects by playing a leadership role in particular areas or by developing statistics portals which operate within the framework of a wider government portal. The fact that there are many strategic, administrative, cultural and technical challenges to overcome in the development of these new initiatives and systems has not prevented significant advances and the experience of statistics offices can help other government bodies in this regard.

Much use has been made of developments in statistical standards and languages such as GEMES and XML to facilitate the exchange of data and metadata. Statistical bodies have grasped the opportunities offered by these advances to produce more efficient exchange mechanisms. National agencies are also supporting initiatives such as the Statistical Data and Metadata Exchange initiative and the IMF Dissemination Standards Bulletin Board. In the case of the DSBB, the role of national co-ordinator is generally assigned to the NSO.

With regard to lowering costs, many statistical offices are increasing the use of data sourced from administrative records in place of direct collection. Apart from achieving budget savings, it also places less of a burden on citizens and businesses which have become increasingly vocal about compliance costs. This option has been facilitated by ICT advances but also requires a commitment to change on the part of both management and staff within the organisation.

Statistical authorities also see different variations of legal, administrative methodological and technical measures to collect and disseminate data in a manner that protects the privacy of respondents by preventing primary and secondary disclosure. These measures are often made possible by ICT improvements, on other occasions ICT advances that provide efficiency gains in disseminating data may require enhanced security measures. Whether by drafting legislation to prevent the abuse of privileged access to data, using the Internet to publicise a confidentiality code, imposing particular restrictions on researchers with special privileges, developing technical barriers such as a statistical firewall as in Sweden or a software programme such as ARGUS, etc. national statistical institutions have risen to the challenge of minimising the risk of disclosure while at the same time making better quality data available.

The advances made by NSOs show that the drive towards e-government requires fundamental changes in the way that public services are delivered. This requires a cultural change within public administrations, with changes to operating practices including better interaction both within and across agencies. These are fundamental challenges that many national statistical offices have met and are continuing to meet successfully. While it is obvious that ICT advances have enabled the collection, compilation and dissemination of statistics to be conducted in a manner that has enhanced their quality, statistical organisations have also had to meet strategic, administrative and cultural challenges. Thus the experiences of national statistical offices can be of benefit to other public sector agencies which may not be as advanced in using e-government as a means of service delivery. This experience can be beneficial not only in an ICT perspective but also with regard to the other challenges that are associated with any major change programme.