IAOS Satellite Meeting on Statistics
for the Information Society
August 30 and 31, 2001, Tokyo, Japan

The State of Information Technology in National Statistical and Census Offices in Asia and the Pacific[1]

SURVO, Ilpo
United Nations Economic and Social Commission for Asia and the Pacific (ESCAP)
Bangkok

Abstract

The utilization of information technology varies considerably in statistical offices in Asia and the Pacific. This is not unexpected considering the diversity of the region and the resources that are available for official statistical systems. This paper compares the state of basic IT infrastructure, web site technology and data capture technologies in selected statistical and census offices. Most of the information has been collected in connection with ESCAP's technical assistance to developing countries.

1. Management challenge

Information technology provides statistical offices means (i) to collect and process data quicker, (ii) to allow users access statistics at their convenience, and (iii) to manage their services efficiently. Although the application of IT is not an end by itself, statistical offices cannot go wrong by dedicating efforts to IT development. To start with, they can adopt solutions that work well in other countries.

The experiences that statistical and census offices have shared in recent ESCAP forums indicate that clearly not enough attention is paid to the management information technology. Attention given and priorities set by the top management is probably the most important success factor in applying IT effectively. When the priority is on improving internal efficiency, statistical data processing and administrative systems tend to be upgraded. Another, not necessarily a conflicting angle is to place importance on the requirements of data users. That tends to lead to improvements in data dissemination vehicles and customer service applications.

2. Basic IT infrastructure

A survey[2] conducted by ESCAP in 1998 confirmed that in-house application development, use of IT across all operations, and universal network connectivity were characteristic features for advanced national statistical offices. Overall, the survey revealed a broad gap between the front-runners and laggards in the effective use and customization of IT applications. NSOs developed applications that suited for their particular needs, but at the same time used off-the-shelf packages for common tasks. Own development was more common for data scrutiny, data editing, data estimation and tabulation phases, whereas data analysis was usually conducted with commercial statistical packages. All responding offices reported a fairly large number of off-the-self software packages, but no significant differences were found in the prevalence of brand names between developed and developing countries (see Annex 1 for most commonly used software).

---

[1] This paper has been issued without formal editing. Mention of any firm, licensed process or product does not imply endorsement by the United Nations. The views expressed are those of the author and do not necessarily reflect the views of the United Nations Secretariat.

[2] The survey was one of the activities of the UNFPA-funded project "Application of New Technology in Population Data Collection, Processing, Dissemination and Presentation". The project web site (see http://www.unescap.org/stat/pop-it/) has an extensive collection of material related to the application of IT in population censuses and surveys in Asia and the Pacific. More detailed results of the 1998 survey can be found in http://www.unescap.org/stat/pop-it/pop-itnl/news_03.htm

In 1998, practically all PCs were connected to network in the national statistical offices of Australia, Japan, Macau, New Zealand, Republic of Korea and Singapore. With the exception of Fiji, the small island developing countries had no LANs installed. The number of their PCs varied from 2 to 32. The largest office without a network (Sri Lanka) had 124 PCs. Local area networking was at an early stage also in Thailand and Turkey, with less than 10 per cent interconnected.

A follow-up questionnaire was sent in June 2001 to the offices that had responded to the 1998 survey. Preliminary results, shown in the shaded columns in Table 1, indicate that during the past three years, the PC, LAN and Internet infrastructure have been upgraded in all responding offices. The increase in the number of computers in use has been remarkable, exceeding 50 per cent in many responding offices. The staff-to-PC ratio has come down because of the increase in the number of PCs; also staff reduction contributed to that in some offices.

Table 1. The number of staff and PCs in selected statistical offices, 1998 and 2001 (sorted by the staff/PC ratio in 1998)

Country/area	Total staff 1998	Total staff 2001	Staff/PC 1998	Staff/PC 2001	PCs in LAN 1998, per cent	PCs in LAN 2001, per cent	Staff change 1998-2000	PC change 1998-2000, per cent
Marshall Islands	6	-	0.5	-	0	-	-	-
New Zealand	729	900	0.8	0.8	100	100	23.5	19.3
Singapore	212	-	0.9	-	100	-	-	-
Australia	2845	3140	0.9	0.8	100	100	10.4	19.3
Japan	1823	1788	0.9	0.8	100	100	-1.9	10.0
Republic of Korea	1281	1568	1.2	0.9	99	100	22.4	57.6
American Samoa	50	-	1.2	-	20	-	-	-
Macao, China	278	-	1.5	-	99	-	-	-
Fiji	70	-	1.8	-	55	-	-	-
Hong Kong, China	1495	1505	1.9	1.0	59	49	0.7	81.5
Lao PDR	50	30	1.9	0.7	85	78	-40.0	76.9
Bhutan	26	-	2.0	-	0	-	-	-
Maldives	5	-	2.5	-	100	-	-	-
Samoa	32	36	2.7	2.0	0	100	12.5	50.0
Philippines	3131	3554	3.3	3.1	26	47	13.5	21.3
Thailand	1254	-	3.7	-	1	-	-	-
Turkey	2741	3079	3.8	2.2	5	71	12.3	91.8
Malaysia	1710	-	4.1	-	57	-	-	-
Indonesia	11942	-	4.3	-	30	-	-	-
Armenia	61	226	4.4	1.5	71	32	270.5	1000.0
Myanmar	311	302	8.9	5.3	29	28	-2.9	62.9
Sri Lanka	1183	-	9.5	-	0	-	-	-
Pakistan, Fed.Bureau of Stat.	1901	-	11.7	-	12	-	-	-
Bangladesh	4428	-	15.5	-	44	-	-	-
Pakistan, Pop.Census Org.	750	-	68.2	-	0	-	-	-

While technologically advanced offices provide email and web connection for all staff, and on every PC, some offices have to manage with a couple of email accounts and a dismal connection speed, which make any web browsing unfeasible (see Table 2). Where a viable Internet connection is missing, field operations have to rely on conventional means of communication, with data being transferred on paper, on diskettes or through direct telephone connections. Web sites, if existing, are likely to be a result of efforts by dedicated individuals using infrastructure outside of their offices.

Table 2. Typical PC configuration and Internet connection in selected NSOs in June 2001

Country	Typical PC processor, MHz	Typical PC RAM, MB	Typical PC hard disk, GB	Type of Internet connection	Speed of Internet connection, kbps	Share of PCs that can send email, %	Share of PCs that can browse web, %
Armenia	100	16	1	Radio modem	4	1.9	6.5
Australia	484*	137*	10*	Frame relay, full duplex	1000	82.1	87.2
Hong Kong, China	333	32	3.2	T1	1544	21.3	21.3
Japan	667	256	15	Dual T1	3000	100.0	15.9
Lao PDR	600	64	10	Cable modem	56	4.3	0.0
Myanmar	166	16	1	None	-	3.5	0.0
New Zealand	200	128	6	Frame relay	2000	90.9	90.9
Philippines	500	64	8	Leased line	64	20.0	20.0
Republic of Korea	700	64	10	T1	2048	100.0	100.0
Samoa	333	64	6	Dial-up	…	38.9	38.9
Turkey	200	32	6	Leased line	128	46.4	46.4

* Weighted average of PCs and notebooks

3. Web sites

A little over half of ESCAP's regional members and associate members, i.e. 32 of 57, had a national statistical web site in June 2001 (see Annex 2). The scope of the sites and material available through them varied considerably, from a limited collection of static pages to an extensive sites providing access to a wealth of statistics and metadata, methodological notes, and corporate and other information.

Although an ordinary visitor cannot see where a particular web server is situated, the physical location affects the speed that users can download information from a web server. Most of the NSO and census web servers appear[3] to be located in the respective capitals (see the last column in Annex 2). Six of the 34 sites are hosted abroad (statistical offices of Azerbaijan, Fiji, Islamic Republic of Iran, Marshall Islands, Federated States of Micronesia, and the Office of the Registrar General of India). Normally, a domestic location gives the best response to local customers, while international customers can download fastest from servers near to the broadband backbone of the Internet.

Server response time

For the purposes of this paper, an experimental test was conducted to compare how fast individual web sites responded to a series of ping requests[4]. The summary results, which are shown in Figure 1, must be interpreted and compared with a caution as they represent response times from one location only, and do not reflect the real performance of each web site[5]. In the Bangkok test, the fasted average response was received from the nearest web site, the National Statistical Office of Thailand, followed by sites located in the United States and ASEAN countries. The slowest responses came generally from the most distant sites in the Eurasian continent. The fast responses from .fj (Fiji) and .fm (Micronesia) are due to their location in Seattle and Honolulu, respectively, which in the Internet topology are advantageous locations in relation to Thailand.

[3] The domain name does not always reveal the physical location of the server. Various public domain tools can be used to trace the route to the server, and to try to identify its location and owner.

[4] Sites that did not respond to ICMP (Internet Control Message Protocol) ping were excluded. The results represent average response times for a total of 1200 individual ping requests, using a 256-kbps frame relay connection from the ESCAP secretariat and a 56 kbps dial-up connection from a home. Response times were recorded over a period of two weeks in May and June 2001.

[5] Network congestion and poor page design can significantly increase download times.

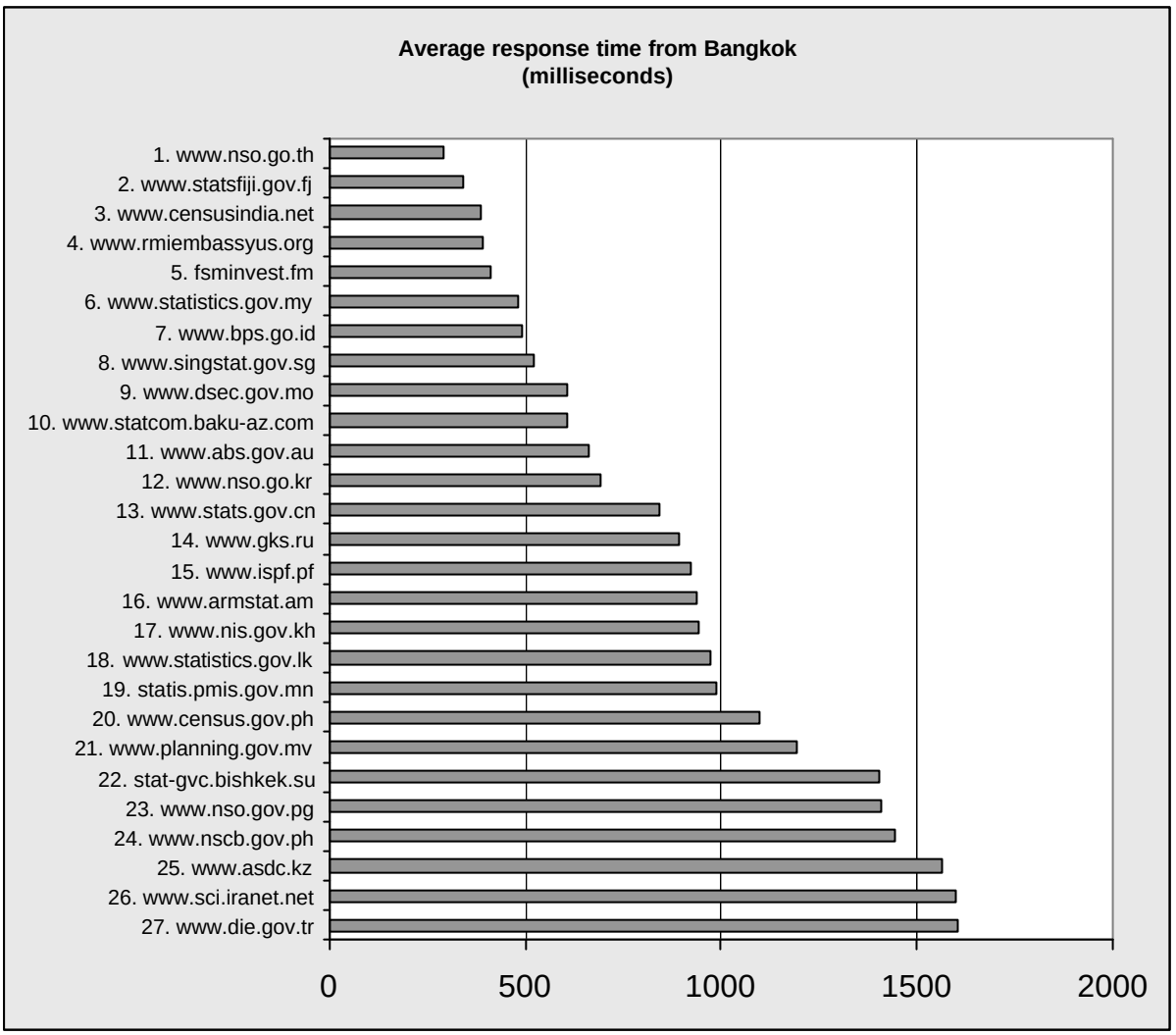

Figure 1. Average response time of statistical web sites to ICMP ping requests from Bangkok.

If the same test were conducted from a third party server located somewhere else, the results would be different. In a sample test, run from servers in the United States and Europe, the web sites of Malaysia and Singapore responded to ping faster than that of Thailand, while a Philippines server took usually a longer time to respond than Thailand. The results reflect available bandwidths between the locations and traffic conditions on the way.

Server platform

The technology of each web site was further investigated through Netcraft's detection service (http://www.netcraft.com). Apache and Microsoft Internet Information Server were by far the most common servers (see Table 3). In addition, Netscape Enterprise and Lotus Domino servers were hosting three and two web sites, respectively. Compared to global server statistics in May 2001, according to which Apache and MS-IIS held 'market shares' of 62 and 21 respectively (see http://www.netcraft.com/survey/), the Microsoft server appears to be relatively more popular among the statistical offices in Asia and the Pacific. The Apache web servers were running on various Unix derivatives, the most popular being Linux (5 servers) and Solaris (4 servers). All fourteen MS-IIS servers, as well as the two Lotus Domino servers, ran on Windows NT4. The Netscape Enterprise servers were on Solaris. Judging from the name of the net block owner, three quarters of statistical and census web servers were maintained externally[6].

[6] Although the server itself is maintained by an outsider, the content is probably produced by the statistical office.

Table 3. Statistical and census web servers by type and location of hosting, June 2001

Web server	Net block owned by		Total	Share, per cent
	NS/census office	Outsider		
Apache 1.3.x	1	14	15	44
Microsoft-IIS/4.0	4	10	14	41
Netscape Enterprise 3.6 – 4.1	2	1	3	9
Lotus Domino 5.0.x	2	-	2	6
Total number of servers	9	25	34	100
Share, per cent	26	74	100	

When outsourcing the maintenance of a web server, statistical offices are likely to pay little attention to what technology their web site would be based on. However, if the server is to be located and managed in-house, attention needs to be paid, among other issues, to the skills required to maintain the server, the server's capability to support content delivery applications, and the compatibility of the technology with the rest of the offices systems (so that sustainable procedures can be established for keeping the content up to date).

4. Data capture technologies

The 1998 ESCAP survey asked how long it had taken (or was expected to take) from the enumeration of the latest population census or housing survey to the tabulation and publishing of final results. In Figure 2, that time is plotted against the size of the census or survey. While the correlation between the size and the total duration of processing was weak, the survey confirmed that in some cases it took (or was estimated to take) several years to compile and publish the results. It should be noted that the results are from censuses and surveys conducted prior to the 2000 round of censuses, involving different technologies, especially in data capture. Some censuses and surveys were already based on optical character recognition (OCR), while others used manual keyboard entry.

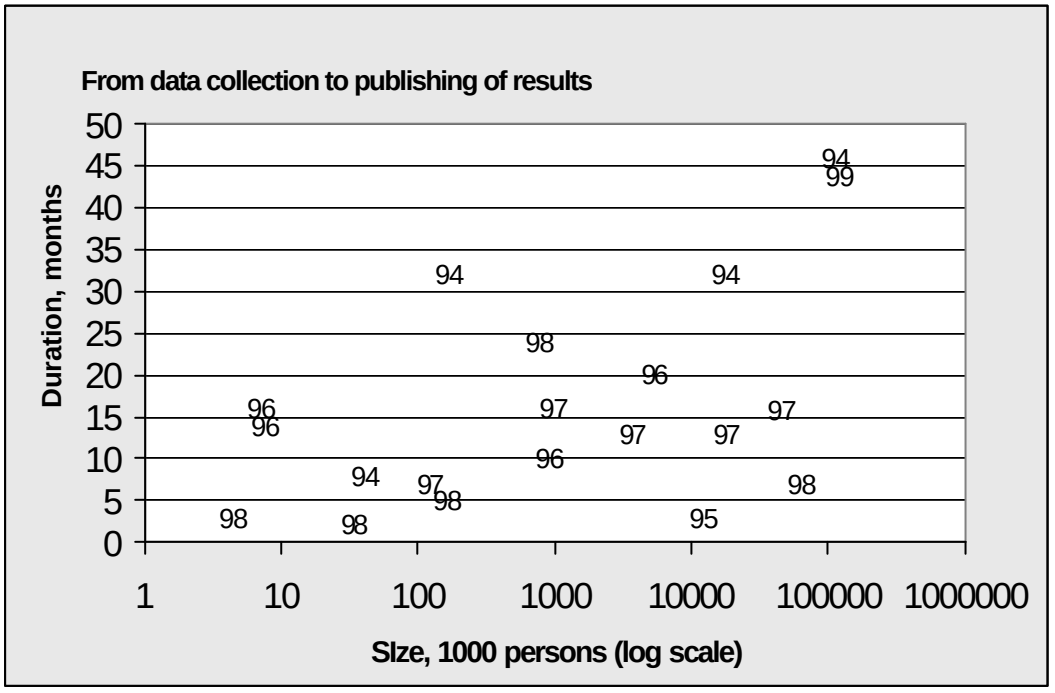

Figure 2. The relationship between the size of population census/survey and the time taken from data collection to publishing of the results

A recent ESCAP workshop[7] gave clear indications that the results from the 2000/2001 round of censuses would be tabulated faster than ever before. The Workshop concluded that data capture through OCR/ICR had become a proven technology that could make significant cost, timeliness and accuracy improvements in census data capture. Several countries that were using OCR or ICR technology for the first time had released preliminary results (based on the whole population) in a matter of a couple of months.

Twelve of the participating 24 offices in the mentioned workshop indicated that their offices still relied on keyboard entry; two used OMR and nine OCR/ICR (see Table 4). The only country to offer the possibility to submit information through the Internet was Singapore, where eventually 15 per cent of the population chose to submit their information through the Internet. Other Singaporeans responded either to computer-aided telephone interviews (CATI) or to person-to-person interviews.

Table 4. Data capture technology in the 2000 round of censuses in selected ESCAP members and associate members

Keyboard entry	OMR	OCR/ICR	Internet+CATI+OCR
Brunei Darussalam	Bangladesh	Australia	Singapore
Cambodia	Pakistan	Bangladesh	
Indonesia		China	
Kiribati		India	
Malaysia		Indonesia	
Mongolia		Macao, China	
Nepal		New Zealand	
Papua New Guinea		Philippines	
Republic of Korea		Thailand	
Samoa			
Sri Lanka			
Viet Nam			

The 'beauty' of optical recognition technologies is that after the questionnaire forms have been scanned into images, they can be split into pieces, question by question or character by character, for recognition in a priority order. Thus, data tabulation and analysis can be started from the most important information and almost immediately after imaging. That is a major advantage over manual keyboard entry, which normally progresses form by form. Handwritten open responses and questions requiring manual coding can be dealt with later as experts and verifiers working on them make progress.

Although the learning curve to master OCR/ICR is relatively steep, the technology has significantly improved the timeliness of the results compared to the previous round of censuses. It has also lowered the total cost of census taking, in some countries by 50 per cent or more. The scanners and recognition software are rather expensive, but the cost can be moderated by using the same technology in several censuses and surveys and by sharing it with other agencies.

5. Lowering cost by using public domain software

Developing countries in the Asia-Pacific region make a good use of free and public domain software. The Integrated Microcomputer Processing System (IMPS) suite continues to be widely used in census and survey planning, data imputation and editing and tabulation. It is expected to be replaced by the Census and Survey Processing System (CSPro), which is an integrated public domain software package for entering, tabulating and mapping census and survey data, combining the user-friendliness of IMPS and the computational power of the Integrated System for Survey Analysis (ISSA).

Statistical and census offices that do not have the resources to procure and use fully-fledged geographic information systems are creating effective maps with light-weight and inexpensive

[7] Workshop on Population Data Analysis, Storage and Dissemination Technologies, 27-30 March 2001, Bangkok, see http://www.unescap.org/stat/pop-it/pop-wdt/pop-wdt.htm.

alternatives. The National Institute of Statistics of Cambodia, for instance, developed a PopMap-based application for a CD-ROM disseminating the results of its 1998 census. Census data could be shown on maps down to commune level, together with line layers for main routes and rivers and point layers for the villages and schools.

6. Future role of IT in national statistical offices

This paper has discussed selected areas of IT, leaving out the heart of statistical data operations, namely data storage, tabulation and statistical analysis. In those areas, the conventional approach, which develops separate relational databases for dedicated purposes and uses separate tools for tabulation and analysis, will be replaced by location-independent access to tabulation and analysis. Statistical offices will move towards data warehousing as related technologies become cheaper and easier to use. Ideally, statistics from censuses and surveys, conducted at different points of time, will be made available through one interface (together with macroeconomic and other statistics). With new technologies available for IP-based networks (internet, intranet, extranet) there is no need to develop systems separately for NSO staff and customers. In other words, once a dynamic online analytical processing system has been developed for the statistical office, selected parts of it can be made available to external users with a marginal development cost.

The world wide web is already the main channel for data dissemination in developed countries, and it will become increasingly popular also in the developing countries, starting from where statistical offices already have an established web presence. The speed, with which online dissemination replaces traditional dissemination in Asia and the Pacific, depends on the capability of customers to utilize such services. However, the missing infrastructure "guarantees" that paper copies will remain the most important dissemination technology in the majority of countries, for the next few years at least.

7. Capacity building by ESCAP

ESCAP, which is the largest of the five regional commissions of the United Nations, works according to priorities and a work plan set by its members. Within its statistics subprogramme, which is implemented under the guidance of the Committee on Statistics, ESCAP has paid increasing attention to capacity building in the area of applying information technology in national statistical systems. Most of the activities are building on utilizing the diversity of the region work for the benefit of statistical offices that are lagging behind in IT infrastructure building, application development and the management of information systems.

Annex 1. Most commonly used software in national statistical offices in 1998

Application	Group 1*	Group 2**	Group 3***
Word processor	Lotus Notes (2); Ami-Pro (1); Word-Pro (1); OASYS [Fujitsu] (1); MS Word (1); Ichitaro8 [Justsystem] (1)	MS Word (3); Hangul 97 (1)	MS Word (21); Borland C^{++} (1); WordPerfect (5); PE2 (1); Acrobat (1); latex (1); WordStar (1)
Spreadsheet	Excel (2); Lotus (1)	Excel (4)	Excel (21); Lotus (6); Quattro Pro (2); Lotus Notes (1); IMPS (1)
Database	Oracle (2); Access (1); SQL Server (1); Sybase (1)	Oracle (2); DB2 (2); FoxPro (1); SAS (1); FAME (1); Natural (1); dBASE (1); Access (1); MS Visual (1)	Access (8); dBASE (7); FoxPro (6); IMPS (3); MS SQL-Server (2); Oracle (1); SAS (1); SPSS (1); Sybase (1); Clarion (1); Clipper (2); Btrieve (1); Paradox (1); RDBS/SQL Version 6 (1)
Graphics	Freelance (1); tailor-made system (1)	Harvard Graphics (2); GDDM (1); MAC (1); PowerPoint (1)	Harvard Graphics (7); Excel (6); PowerPoint (4); Corel Draw (3); Quattro Pro (1); Paint Brush (1); MS Publisher (1); Lotus (1); Freelance (1); PageMaker (1); PhotoShop (1)
Analysis	SAS (3); tailor-made system (1); Access (1)	SAS (4); SPSS (1); FAME (1)	SPSS (11); SAS (9); IMPS (2); STATA (2); ISP (1); SPLUS (1); MortPak (1) ; PERT (1); MS Word (1); TSP (1); Q5 (1); Excel (1)
Mapping/GIS	MapInfo (2); Tailor-made system (1); ArcInfo (1)	ArcInfo (1)	MapInfo (5); ArcInfo (4); GIS (2); AutoCAD (2); PopMap (2); ERDAS (1); ArcView (1); Atlas (1)
Population projection	SAS (1); SuperCROSS (1); Excel (1); MapInfo (1); Lotus (1); tailor-made system (1)	Tailor-made system (1); POP SYN (1); QBasic (1); Lotus (1); dBASE (1)	People (13); Mortpak (3); DEMPROJ (2); Labour (1); Fortran (1); Fivsin (1); PAS (1); Workers (1); Excel (1); Basic (1); IMPS (1)
Other commonly used software	SQL Windows (2); Outlook (1); PC3270 Emulator (1); ACOS ETOSA Emulator (1); PowerPoint (1); Blaise (1); SSA-Names3 (1); SuperCROSS (1); Visual Basic (1); Novell V4 (1)	Lotus Notes (1); Mcafee Anti-virus (1); X-11-Arima (1); WordPerfect (1)	IMPS (3) Visual Basic (3); COBOL (3); Lotus (1); AmiPro (1); EASWESPOP (1); ISSA (1); Access (1); Paradox (1); Delphi (1); Redua (1); Pascal (1); Desktop Publishing Software (1); RPG (1); PowerPoint (1); Fortran (1); C Programming (1)

* Developed members of ESCAP: Australia, Japan and New Zealand

** Members and associate members of ESCAP which are either newly industrializing economies or developing members of OECD: Hong Kong, China; Republic of Korea; Singapore and Turkey

*** Developing members and associate members of ESCAP: American Samoa ; Armenia; Bangladesh; Bhutan; China; Fiji; Guam; Indonesia; Islamic Republic of Iran; Lao People's Democratic Republic; Macao, China; Malaysia; Maldives; Marshall Islands; Myanmar; Pakistan; Philippines; Samoa; Sri Lanka; Thailand and Turkmenistan

Annex 2. Web sites of statistical and census offices in the ESCAP region, June 2001

Country/area		Organization	URL	Location of web server
1	Armenia	State Registry and Analysis of the Republic of Armenia	http://www.armstat.am/	Yerewan
2	Australia	Australian Bureau of Statistics (ABS)	http://www.abs.gov.au	Sydney
3	Azerbaijan	State Statistical Committee	http://www.statcom.baku-az.com/	Lithuania
4	Cambodia	National Institute of Statistics	http://www.nis.gov.kh	Cambodia
5	China	National Bureau of Statistics	http://www.stats.gov.cn	Beijing
6	Fiji	Bureau of Statistics	http://www.statsfiji.gov.fj	Seattle
7	French Polynesia	Institut Territorial de la Statistique	http://www.ispf.pf	French Polynesia
8	Guam	Department of Commerce	http://www.admin.gov.gu/commerce/	Sydney
9	Hong Kong, China	Census and Statistics Department	http://www.info.gov.hk/censtatd/	Hong Kong
10	India	Ministry of Statistics and Programme Implementation	http://www.nic.in/stat/	New Delhi
	India	Office of the Registrar General	http://www.censusindia.net/	Ottawa
11	Indonesia	BPS - Statistics Indonesia	http://www.bps.go.id/	Jakarta
12	Islamic Republic of Iran	Head, Statistical Centre of Iran	http://www.sci.iranet.net	Kuwait
13	Japan	Statistics Bureau	http://www.stat.go.jp/english/1.htm	Tokyo
14	Kazakhstan	National Statistical Agency	http://www.asdc.kz/kazstat/	Kazakstan
15	Kyrgyzstan	National Statistical Committee	http://stat-gvc.bishkek.su/	Bishkek
16	Macao, China	Direccao de Servicos de Estatistica e Censos	http://www.dsec.gov.mo/	Macao
17	Malaysia	Department of Statistics Malaysia	http://www.statistics.gov.my	Kuala Lumpur
18	Maldives	Ministry of Planning and National Development	http://www.planning.gov.mv/index2.htm	Maldives
19	Marshall Islands	Office of Planning and Statistics	http://www.rmiembassyus.org/stats.html	Sudbury (US East Coast)
20	Federated States of Micronesia	Office of Planning and Statistics	http://fsminvest.fm/Statistics/	Honolulu
21	Mongolia	State Statistical Office of Mongolia	http://statis.pmis.gov.mn	Mongolia
22	New Caledonia	Institut Territorial de la Statistique et des Etudes Economiques	http://www.itsee.nc/	New Caledonia
23	New Zealand	Statistics New Zealand	http://www.stats.govt.nz/	Auckland
24	Papua New Guinea	National Statistical Office	http://www.nso.gov.pg	Port Moresby
25	Philippines	National Statistical Coordination Board	http://www.nscb.gov.ph	Manila
	Philippines	National Statistical Office	http://www.census.gov.ph/	Manila
26	Republic of Korea	National Statistical Office	http://www.nso.go.kr/	Seoul
27	Russian Federation	Russian Statistical Agency (Goskomstat)	http://www.gks.ru/	Moscow
28	Singapore	Department of Statistics, Ministry of Trade and Industry	http://www.singstat.gov.sg	Singapore
29	Sri Lanka	Department of Census and Statistics	http://www.statistics.gov.lk	Sri Lanka
30	Thailand	National Statistical Office	http://www.nso.go.th	Bangkok
31	Turkey	State Institute of Statistics	http://www.die.gov.tr/ENGLISH/	Ankara
32	Uzbekistan	State Department of Statistics	http://www.gov.uz/mms000.html	Uzbekistan

Statistical offices without a web site (25):

Afghanistan, American Samoa, Bangladesh, Bhutan, Brunei Darussalam, Cook Islands, Democratic People's Republic of Korea, Georgia, Kiribati, Lao People's Democratic Republic, Myanmar, Nauru, Nepal, Niue, Northern Mariana Islands, Pakistan, Palau, Samoa, Solomon Islands, Tajikistan, Tonga, Turkmenistan, Tuvalu, Vanuatu, Viet Nam.

Source: A list of national statistical offices maintained by ESCAP, http://www.unescap.org/stat/nsos.htm. The physical location of each web server was detected with IP tools 'traceroute' and 'whois'.