STATISTICAL DATA PROCESSING IN DEVELOPING COUNTRIES: APPLICATION OF EMERGING TECHNOLOGY [1, 2]

by

George Sadowsky[3]

I. INTRODUCTION

All countries of the world are faced with a substantial number of major challenges. Among them are the provision for their inhabitants of good health, adequate education, opportunities for advancement, adequate housing, employment, sufficient income to meet material needs, a sense of personal security within the law, and a sense of security as a nation. While individual countries may disagree about how to achieve these goals, there is agreement about what they should be in a general sense.

In making progress toward achieving these goals, many countries currently make intensive use of information about the state of their societies and rely upon information technology to exploit it. National statistics are important. They provide a measure of the economic and social state of an economy, a benchmark of where an economy is today. They contain an inventory of the productive capacity of an economy, the public and private investment in assets of many types, the ability of the economy to provide employment for workers, the degree and nature of interdependence with other countries through foreign trade, and the economic performance of the country relative to the rest of the world. Statistics over time measure changes in the economic and social status of the population and provide evidence regarding the relative progress of development programs and success from policy measures undertaken by the government as well as the effects of external events upon the country. The existence of adequate, accurate and timely data is a vital element for supporting development planning, implementation, and program monitoring.

The primary mission of national statistical services is to provide relevant, reliable, and timely information for a wide range of uses in making decisions and in assessing the state of and studying changes in society. The production of official statistics requires many skills and operational activities, among them the exploitation of computer technology.

---

[1] Parts of this paper were originally prepared as a discussion paper for the Interregional Workshop on Statistical Data Processing and Data Bases, sponsored by the United Nations Department of Technical Co-operation for Development in co-operation with the United Nations Statistical Office and the Economic Commission for Europe, which was held in Geneva during 30 May-3 June 1988.

[2] Preparation of this paper has been assisted by helpful conversations with and advice from members of the staffs of the Statistical Institute for Asia and the Pacific, the Economic and Social Commissions for Asia and the Pacific and for Africa, the Statistical Offices of Malaysia and Kenya, the Food and Agriculture Organization and the United Nations Statistical Office. They are not responsible for the information contained in this paper or for the conclusions and recommendations presented.

[3] The author is currently Director of Academic Computing and Network Services at Northwestern University in Evanston, Illinois, U.S.A., and was formerly a Technical Adviser in Computer Methods for the United Nations Statistical Office and Department of Technical Co-operation for Development. The views expressed herein are those of the author and do not necessarily reflect either those of the United Nations or of Northwestern University.

Social and economic activities today are sufficiently complex that computing technology is an essential component of most if not all statistical activities.

Developing countries generally lag in the use of computer technology to process their statistical data. Yet their needs are just as great; they need to achieve a viable statistical data processing capability if they are to provide, on a continuous and sustained basis, the essential statistical information needed for development planning and administration in their countries. This disparity between capabilities and demands applies not only to statistics but also generally to many sectoral areas in developing countries, and much of the discussion that follows is relevant for the development of an acceptable data processing infrastructure throughout the public sector, independent of area of application.

A major difficulty in discussing the application of computing technology in developing countries is that the countries as a group are quite heterogeneous. Developing countries differ widely in the extent to which they have introduced computer technology, and the extent to which the necessary infrastructure exists for exploiting the technology. As a consequence, for almost any statement one can make on this subject, there is highly likely to be a group of countries for which it may be either not meaningful or not applicable.

II. OVERVIEW OF STATISTICAL DATA PROCESSING AND DATA BASES IN THE DEVELOPING COUNTRIES

Data processing technology has long been an essential component of the operations of national statistical offices. The urgent need for timely results from United States decennial population censuses led the U. S. Bureau of the Census in 1881 to commission Herman Hollerith to design and construct data processing machinery capable of tabulating census results rapidly and automatically. In the intervening four decades, electronic data processing (computing) equipment has become affordable and usable by most if not all national statistical offices in the world.

Statistical data processing, as understood within the context of national statistical offices, covers many fields of economic and social activity. Typically, data of active interest include data describing individuals and households, industrial production, agricultural production and land use, energy production and consumption, indicators of social and environmental status, international trade, construction, health, education and human resources, labor force participation, tourism and immigration, finance, balance of payments, transport, communications, distribution, and national accounts.

III. SPECIAL PROBLEMS IMPEDING DEVELOPMENT IN STATISTICAL DATA PROCESSING AND DATA BANKS

Developing countries are faced with a set of problems that impede their ability to take good advantage of computing technology. Key factors that affect the rate of absorption and utilization of computer technology by a country include both internal and external factors. Internal factors include the existing physical and human resources infrastructure of the country, and its state of development and rate of growth. External factors include the willingness of suppliers to do business in the country and the availability and amount of international assistance directed toward transfer of this technology.

Physical infrastructure. The degree to which physical infrastructure is important is now changing in large part due to the introduction of microcomputer technology. Mainframe and minicomputer systems require a reliable supply of electrical power and reasonably stable environmental conditions such as temperature and humidity levels in order to operate reliably.

Almost all developing countries now have the necessary physical infrastructure to support operations of one or more types of computer systems, at least in one or more principal cities. Where electrical power is unreliable, power regulators and/or standalone generating capacity may be required for larger computer systems, making use of such systems more expensive than it otherwise would be.

Transportation and communications links to suppliers of hardware and software resources are also required for effective use of computers.

Lack of adequate physical infrastructure is rarely a major problem today, although remedying deficiencies in it will generally increase the costs of providing computing capability to a country. As overall economic development occurs, this situation should improve at a moderate rate.

Human resources infrastructure. The skills required to exploit computing technology effectively depend upon minimal levels of competence within the country in fields such as engineering and mathematics. The level of literacy among the most educated people may be an important factor. In some countries, the ability to work in a foreign language used by suppliers of computing equipment and software may be an important determinant of success. Aspects of the cultural milieu within the country may also be important. Responsiveness to educational opportunity, the strength of the work ethic, and attitudes and policies toward achievement, employment, production and productivity are all important in the successful transfer of this technology.

In general, a sufficient human resources infrastructure exists in most if not all developing countries to be able to make use of computer based systems. However, it is often thin, and there is substantial competition for the scarce more talented individuals within both the public and the private sector as well as between them. Emigration to better labor markets (the "brain drain") causes a depletion of the resources necessary to exploit this technology; in countries initially having a limited set of human resources with which to work, such a drain may cause data processing activities to break down completely for periods of time. The lack of priority given to computing in many developing countries, along with their inability to develop personnel policies adequate to attract and hold qualified specialists in the public sector is well known and widespread. In many countries, no solution is in sight for this problem, and a new approach may be needed.

Size of market. Before microcomputers were developed, the current and potential size of the market for computer systems was an important determinant of the potential profitability of entry into the country by suppliers. Countries that had limited markets generally received attention from one supplier; some countries had no suppliers of computing equipment and had to depend upon sales, support, and service to come from another country. Within countries having limited markets for computing equipment, the price of conventional equipment supplied in the past has been quite high [4] Countries with

[4] Such high prices for conventional equipment have in the past been cited as examples of exploitative behavior, typically by multinational countries in their dealings with developing country economies. While

relatively high rates of growth attracted more than one supplier and generally enjoyed the benefits of a more active domestic computer industry, including product variety, more competitive markets, and greater responsiveness of the market to demands for specific products and services.

The introduction of microcomputers into developing countries has modified the situation drastically. Microcomputer systems are in general relatively inexpensive, easily shipped, and easy to maintain.

The most serious constraint upon computer hardware acquisition have been and continues to be the necessity for and availability of computer hardware maintenance services.

Information poverty. People living in developed countries live in an information rich environment. A rich variety of information is available daily from newspapers, magazines, books, government publications, technical and scholarly journals, radio, and television. Information is further available from friends and colleagues, no matter whether they are local or distant. Reliable telephone systems provide relatively economical voice communication between most or all parts of the country. Emerging electronic mail systems connect professionals of many kinds together in an effective manner for passing written messages; such systems are increasing to cover larger subscriber bases and larger geographic coverage.

Information poverty is one of the more important and insidious obstacles in developing countries to effective exploitation of information processing and other types of technology. The lack of adequate information regarding developments in other countries and other environments is often not noticed; in the absence of new information, old techniques and procedures are continued without conscious knowledge of alternatives.[5] And, while developing countries may not be hurt in an absolute sense by lack of information, they are certainly negatively affected by any relative measure.

The issue underlying information poverty is how to overcome it to provide a sufficient flow of information into developing country environments so that individuals and governments can make good choices from among realistic alternatives. There are a number of approaches that could assist countries in obtaining a richer and more up-to-date flow of relevant information for statistical data processing. They include: (1) stronger partnerships with national universities and similar organizations; (2) greater exploitation of both short and long term visits by experts from other countries; (3) effective sharing and exploitation

---

such exploitative behavior cannot be denied, a sufficient reason for higher than average prices at that time was the high level of support costs incurred by the supplier in servicing a limited market. There are economies of scale in supporting mainframe computers in compact geographic areas, and markets in many countries were too limited to realize such economies. Limited markets also were conducive to the perpetuation of monopoly power, and therefore monopoly pricing, by sole suppliers in a country.

[5] Readers who wear eyeglasses and whose vision is changing may feel some sympathy with this condition. Deterioration of corrected vision is generally sufficiently slow that it is not noticed *until* new eyeglasses with improved correction are first worn. Often the effect of first using such new eyeglasses is startling; objects and details snap into sharp focus and the wearer is amazed that such improvement is possible. With respect to information, the effect is similar; it is often difficult to understand the losses due to not having relevant information until you have it, and only then is it possible to assess the opportunity cost of not having had it.

of locally available technical skills and knowledge; (4) establishment of an informal statistical data processing newsletter; and (5) recognition of the importance of release time for continued on-the-job learning.

Financial poverty. Most developing countries are financially poor relative to developed countries. They suffer from lack of both financial assets and national income. Their economies are subject to substantial performance fluctuations due to factors beyond their immediate control. Although the rapid and sustained decline in the cost of information technology has been of substantial help in exploiting information technology over time, often sufficient resources cannot possibly be allocated for its acquisition given the many other major pressing needs of developing countries.[6]

Technology transfer assistance. Many developing countries are benefitting from direct assistance in transferring computer technology to themselves. Involvement with private sector firms in developed countries can have significant benefits within an appropriate context; policies promoting domestic investment and taxation and profit repatriation incentives can encourage firms to enter local markets and provide benefits to the country. Private foreign investment in high technology fields often brings with it significant flows of information and training opportunities.

IV. PERSONNEL DEVELOPMENT AND TRAINING

One of the most fundamental obstacles to the production of timely and reliable statistics is the availability of adequate levels and types of data processing skills within statistical offices.

The shortage of professional staff having data processing skills is not limited to developing countries. Since the inception of the industry about 40 years ago, the use of computers has grown rapidly and there has been a persistent shortage of trained manpower in most fields of application and in most countries. The premium in wage rates in developed countries for trained data processing personnel reflect the shortage of such staff. Less visible (especially to observers located outside such countries) but nevertheless equally telling is the rapid job turnover of people having computing skills.[7]

---

[6] Poverty is to some extent relative, and even in developed countries many offices and organizations dependent upon computer technology believe that they are allocated insufficient resources to perform their tasks adequately. The response in developed countries is often to intensify the search for more effective ways of obtaining more output from limited inputs; i.e. to find ways of expanding the production possibility set. Most developing countries are worse off in two ways; they have substantially less resources in an absolute sense than do the developed countries, and the other major needs of developing countries, such as education and health care, are likely to be more urgent than in developed countries. Bilateral and multilateral assistance is available to ameliorate the financial situation to some extent, but it rarely can close the gap in a very substantial way.

[7] Rapid job turnover has been common in data processing labor markets in developed economies for the last 40 years, ever since the inception of electronic data processing as a recognized career path. Although the excess demand for computing skills identified by this market condition has been partially alleviated by training efforts, a greater part of the solution has come from the development and use of computer software that has expanded the set of people who can use computer effectively to a set well beyond only computer professionals. Since such software is relatively easily exportable between countries, developing countries may want to depend more upon such an approach rather than upon one which places primary emphasis upon training pure computer professionals.

The effect of these shortages upon the developing countries is more severe than it in developed countries. In developing countries, there is often no critical mass of specialists to begin with, so that filling a vacancy may require a long time, even to the point of having to train a replacement from the beginning. Rapid turnover may be tolerable when rapid replacement can be anticipated. However, the absence of any replacement may be sufficient to stall progress on a project indefinitely.

An effective training strategy in data processing skills must be regarded as an essential condition for both survival and development for a national statistical office. It must be accompanied by a strategy that assures retention of the skills within the office to as great an extent as possible, else the training will not directly benefit the purpose for which it was intended. Among the important issues are: (1) who is to be trained, in what subjects and to what depth; (2) who can best provide the training required; (3) what should be the relationship between conceptual training and detailed operational training; (4) how should the training be distributed between local and foreign locations; (5) how can indigenous training capacity be built up and/or utilized; and (6) how can changes in the organizational structure assist in ensuring continuity of the required data processing skills being available to the statistical office?

V. APPLICATION OF EMERGING TECHNOLOGIES IN THE DEVELOPING COUNTRIES

Developing countries have historically lagged developed countries in utilizing computer and related information technology products for statistical data processing. Until recently, the support requirements for mainframe and minicomputer systems were significant, and computer suppliers would generally supply products that were somewhat older and more well understood than the leading edge products being introduced in developed countries at the same time. Such a strategy worked well in part also for the developing country; the equipment and software supplied was well understood and more stable than when it was first introduced. Suppliers could afford to offer and support such systems far from their geographic areas of technical expertise because the systems were well understood and trouble free relative to newer models.

With the rapid development of the microcomputer industry, the minimum practical lag between the introduction of current technology in developed and developing countries has been compressed to a matter of months instead of years.[8] This lag no longer depends very much upon the ability of the original manufacturer to support its products in a specific geographic region, since support can now often be delivered either within the country or with the assistance of a third party, perhaps an international aid agency, either locally or remotely. Instead, the lag depends more upon the ability of trained personnel in the country who are able to exploit the microcomputer systems and train others to do so also. This function can initially be provided by an expert from outside the country and later, as the results of training accumulate, by local expertise. For the first time, emerging

---

[8] This argument is of course valid only if the funds exist to acquire the equipment in the first place. Assuming however that funds *are* available, the range of appropriate information technology products that can now be supported within developing countries is relatively larger than it has been in the past because the newer technology requires less dependence upon the original supplier of the equipment

computing technology can be placed at the disposal of developing countries almost as soon as it is readily available and understood in the developed countries.[9]

Microcomputers are small, relatively inexpensive, counterparts to previous generations of mainframe and minicomputer systems that dominated computing in past decades. In terms of processing power and primary storage, they now equal or surpass capabilities that existed in developing countries a decade ago. Their capacity for handling mass storage, especially inexpensive sequential storage, is less well developed, but such capabilities are increasing over time.

Microcomputer technology is appropriate for almost all phases of statistical data processing. In addition, its personal interactive nature and the available software make it in many respects a better computer platform on which to base statistical data processing activities. Microcomputers can be used effectively for data collection, editing, tabulation, analysis, publication, control, and management. This technology provides the potential for a fundamental breakthrough in the effectiveness of statistical data processing operations. **Statistical offices in developing countries should base their future statistical data processing strategy upon the use of microcomputer technology at much of the working level of the office**.

Microcomputers, however, are not a panacea for data processing problems, just as the older mainframes did not solve all problems. Among the issues raised by microcomputers include: (1) adaptation of microcomputer technology to classical steps of statistical data processing; (2) additional activities tractable with microcomputer technology; (3) new responsibilities assumed by microcomputer users; (4) issues of data control and ownership; (5) revised selection criteria for computer processing environments; and (6) exploitation of a broader spectrum of benefits from microcomputer technology.

Adaptation of microcomputer technology to classical procedures in statistical data processing. The movement of the classical activities of statistical data processing -- data collection, editing, tabulation and analysis -- to microcomputers has taken place in several ways.

First, established programs that previously executed in mainframe and minicomputer environments have been moved without significant changes to microcomputers. Examples of such programs are CENTS4, CONCOR, TPL, SPSS, BMDP, PSTAT, SAS, and the population projection program sets from both the U.S. Bureau of the Census and the National Academy of Sciences.[10] These migrations are possible because the

---

[9] Progress in the computer base for statistical data processing depends upon advances in hardware and software, the rate of technical innovation, and the rapidity with with these advances are translated into marketable products at affordable prices. The rate at which the technology is exploited within any given environment is another matter, and reflects the status of the local infrastructure, especially the human resources component, and other conditions between countries.

[10] It is interesting to note different paths that were taken by these programs en route to a microcomputer environment. CENTS4 and CONCOR were originally written to fit into small mainframe computers, and were rather easily adapted to microcomputer environments. SPSS and PSTAT had their origins in the 1960's and were first implemented on the largest mainframes available at that time; the adaptation to microcomputers was more difficult because of the substantial amount of code that had to be converted. One of the most complex programs, TPL, written by the U.S. Bureau of Labor Statistics was not converted; the system was entirely rewritten for a microcomputer environment by QQQ Software Inc. of Arlington, Va.

microcomputer system configurations available to developers today often are at least the equal of the mainframe environments in which the programs were developed originally. Limitations in the microcomputer versions are generally related to mass storage limitations in microcomputers generally[11] and possibly the speed of scientific calculations in microcomputers not having specialized hardware to execute them efficiently.[12]

In addition, new programs not specifically developed for use in statistical data processing have been adapted to perform such work. The most prevalent example appears to be the use of Lotus 1-2-3 to process aggregate data and surveys of modest size and to provide a framework for simply structured statistical data banks. Lotus presents a spreadsheet interface to its user which is more suited for accounting than statistical data processing operations, yet it has high value for work in statistical offices.[13] Similarly, some simple data base management programs such as DBASE 3 have been used to implement both numeric storage and processing of small data collections as well as administrative operations that were either performed with text editors or special programs in previous computer environments or were performed manually.

Furthermore, access to microcomputers is so widespread that many statisticians and professionals in related groups have written simple yet effective interactive programs to address specialized requirements. The ESCAP program for national population projections is an example such a program that has been documented and published by ESCAP. Many such programs are ad hoc and are not widely publicized, but they are useful nonetheless. The availability and ease of use inherent in many microcomputer user interfaces makes these developments possible.

---

by one of TPL's primary developers. There are now few mainframe statistical systems of general use that have not either begun or completed a migration in some form to a microcomputer based implementation.

[11] Within the developed countries, current product development work and actual products brought to market indicate that the mass storage restrictions are being alleviated. The marketing problem in doing so has been to provide individual users with adequate mass storage on a personal basis. The comparison for statistical offices is between microcomputers with large mass storage and an earlier mainframe with very substantial mass storage, but on a *shared* basis. While this issue is not yet resolved, it may be that statistical offices will need or want to rely upon large disk and file servers connected to individual microcomputers through local area networks.

[12] Cost considerations are interesting in this area. Within the last 10 years, additional such hardware, generally called *floating point* hardware, could be added to larger computers for about US$5,000 or more. The equivalent hardware in a microcomputer concept is a floating point chip; these currently cost about US$75-$300.

[13] An interesting related development has been the recent emergence of programs which make use of the spreadsheet metaphor within the context of official statistics represented in tabular form. A good example is provided by PRTAB from Prospect Research in New Haven, Conn. Such a model allows data to be represented in spreadsheet or tabular form, with the advantages of being able to manipulate the table in standard ways; it *also* provides the additional attributes and superstructure that characterizes official statistical tabulations: footnotes attached to rows, columns, and entries, grouped headings; an expanded notion of units of measure; non-cardinal data classifications for categories such as missing, insignificant, and not applicable, with the spreadsheet algebra extended to cover such categories. While the extent to which such an approach is useful remains to be determined, it is a promising development for the tabular processing and storage of certain common types of statistical entities.

Finally, for data entry procedures that were already being executed on primitive microcomputers using a combination of physically oriented controls and arcane languages[14], the move to more general and powerful microcomputer environments must have been very welcome. A new generation of data entry programs, typified by Entrypoint and Rode-PC, provide considerable power, integration of data entry and editing, and use of multi-purpose hardware that can also be used for most other functions in the office. Portable computers allow the concrete possibility of substantial decentralization of the data collection activity up to the actual point of collection.[15]

Additional activities tractable with microcomputer technology. Advances in technology help organizations to perform current activities better, where better performance may mean faster, or more inexpensively, or more efficiently, or some combination of these methods of extracting savings. In the long run, the more important benefits of technical progress are derived from new activities that such progress makes feasible; these are activities that either could not possibly have been performed before the breakthrough or could have been performed either at such exorbitant cost or under such unacceptable conditions that not attempt would have been made to do so.

Inexpensive computing in the form of microcomputer technology allows additional activities to be performed that were not feasible in the past. The possibility of decentralized data entry, including use of portable machines, mentioned above is one such possibility. Two additional activities, microcomputer based publication and data distribution, deserve discussion. Yet another possibility, the evolution of geographic information systems, is discussed in the next section.

The traditional form of dissemination of statistical results has been through publications. Traditionally these have been in printed form, although microform substitutes have been used in developed countries and for large publications. Although newer and considerably more powerful forms of dissemination are emerging, traditional statistical publication will continue, albeit in modified and possibly diminished form, for many years.

Statistical publications initially were produced using master copies in typewritten or typeset form. Both forms of production require transcription of the data from either manual worksheets or computer output, increasing the possibility of transcription errors. More recent methods include preparing camera ready copy directly on the computer[16] and

---

[14] Classical data entry devices consisted of card punch machines, later replaced by diskette recording devices such as the IBM 3742 and the IBM 5280. The 3742 did little more than modify the recording medium adding limited programmability. The IBM 5280 was programmable in a non-standard dialect of RPG which was not an easy language to use, requiring professional programmers.

[15] There is some controversy regarding the degree to which the computer should itself be interjected directly into the interview or primary data collection process; this issue is not discussed here. However, there are clearly a significant number of instances at present for which it would be beneficial to couple the data collection and recording processes. The benefit has three possible dimensions: (1) increased speed of recording and therefore earlier completion of the activity and dissemination of results; (2) cost savings due to elimination of intermediate data recording steps; and (3) increased quality of data by using the same person to collect and record data. The last benefit is the most intangible and least measurable quantitatively, but may be the most important in some activities.

[16] The preparation of camera ready copy directly on the computer had mixed results. On the one hand, it eliminated the data transcription process, which improved accuracy of the publications. On the other hand, it substituted a rough and relatively unattractive single font and single point size monospace format for

preparing from a computer in machine readable form that is then read as input by photocomposition equipment. Both methods eliminate the data transcription step; however, both methods are relatively inflexible and have other disadvantages. Direct photography of computer output sheets places a formatting burden upon statistical data processing programs that they are generally ill equipped to handle, while a computer supported linkage to photocomposition equipment must take into account the idiosyncratic nature of specific photocomposition equipment.

Microcomputers have made possible a new alternative, which is often referred to as *desktop publishing.* The technology depends upon two elements: (1) inexpensive laser printing technology which has recently been developed that provides reasonably crisp images of facsimiles of typeset pages; and (2) the emergence of page description languages, which allow pages to be composed in a fairly flexible and interactive manner.[17] Current standard laser printers produce page images at a resolution of 300 dots per inch, but printers with finer resolution exist and are decreasing in price as the market for such devices grows. Combined with other image manipulation tools on microcomputers, the user has the flexibility of using page layout software that is already moderately powerful along with additional tools for publication layout.[18] Further, the content of the pages can be directly transferred from spreadsheets and data bases, as well as being generated by special purpose programs.

Microcomputer based publishing also allows both the creation and the publication of graphical material in an integrated form. Graphs are readily produced on microcomputers, either using procedures coupled to spreadsheet programs or independent programs. Using computers which have a common bit mapped base for both text and graphics such as the Macintosh, integration of text, tables, and graphics is straightforward, as is the production of high quality pages containing all three types using desktop publishing techniques. The incorporation of graphics into statistical publications often adds to their attractiveness, readability, and ability to convey to the reader the meaning of the statistics reported.

A particularly interesting and powerful aspect of using computers which are based upon bit mapped graphics output is the capability to represent non-Latin national character sets. For example, using the Macintosh, it is relatively easy to define character set fonts for many national character sets and then intersperse them easily and flexibly within tables and mix them with Latin characters and variations thereof. This capability is an integral part of the computer's architecture, in particular: (1) the integration of text and graphics at the most fundamental level through use of bit mapping for both screen and printed output; and (2)

---

previously typeset pages which were easier to read and were more economical in their use of page space. Many users may have perceived the change as negative because of the degradation of form , even though the substance was improved. Form does have importance; in the past, one may have had to make a tradeoff between them, but with the advent of newer techniques they may no longer be substitutes.

[17] There are several heavily used page description languages available in the market today. The most widely used and a likely long term standard is Postscript, produced by Adobe Systems.

[18] The low cost leader in desktop publishing is Apple Computer using its Macintosh, although IBM has recently begun to offer such products with its PS/2 line of systems. Apple's lead in both desktop publishing and image manipulation tools is sufficiently great that it can be considered the vendor of choice wherever Apple support is available. Microcomputer data files can be transferred between MS/DOS and Macintosh computers with relative ease, so that the relative advantages of the Macintosh software environment can be exploited even if the majority of processing is performed in an MS/DOS environment.

the ability to separate and redefine the size and shape of the graphic symbol associated with each key and combination of keys on the keyboard.[19] Publications prepared with such systems can therefore respond well to local needs when two or more different languages and/or character sets would assist users.

Desktop publishing and image manipulation tools implemented on microcomputers are generally easy to use, and require no specialized computer training. Responsibility for the production of publications using such tools can therefore be assigned to statisticians and their colleagues, with appropriate levels of recommendations and consulting assistance from a central data processing staff. Thus, there are no technical impediments that prevent the model developed above of the statistician assuming responsibility for the entire operational cycle of data processing to being adopted for the publications stage also. Desktop publishing technology is powerful and relevant for, *inter alia*, national statistical offices, and deserves thorough investigation and adoption.

From the early days of use of computers, machine readable dissemination and interchange of data has been an attractive goal. For many purposes, it has been reached but not without substantial complications caused by the variety of incompatible formats introduced and supported by the various mainframe and minicomputer manufacturers. Because of the manner in which it has occurred, the adoption of microcomputers by national statistical offices may have provided an effective *ad hoc* partial solution to the data interchange problem.

Statistical offices have converged almost uniformly upon the use of the MS/DOS[20] family of operating systems for their microcomputers. Within that computing environment, the most widely accepted diskette format, that can be read by almost every such machine and which is the default format for software distribution, is the use of double sided, double density diskettes holding approximately 360 KB of information. Text files[21] written on this physical format standard can be interchanged between almost any two machines. Lotus worksheet files can be interchanged between any two systems having copies of Lotus 1-2-3[22]. **The MS/DOS 360 KB format, together with the use of text**

---

[19] The 1-1 correspondence between keys and key combinations and graphic symbols places a limit on the number of characters that can be represented in this manner; Japanese and Chinese character sets and perhaps a few others cannot be handled in this manner. However, there exist systems more sophisticated than the Macintosh, also using bit mapped screens as user interfaces, which accommodate such character sets in a multi-font environment. Advanced work in this area is being performed by Joseph Becker at Xerox PARC in Palo Alto, California. The utility of the overall approach can be seen by noting that there now exist commercial fonts for the Macintosh, available at relatively low prices, for many non-Latin character sets.

[20] The MS/DOS family of operating systems includes the first member of the family, PC/DOS, for IBM microcomputers. However, within statistical offices and in both developing and developed countries, clones dominate the marketplace. The use of the term clone is not pejorative in this context; although there are clones that have significant incompatibilities with the *de facto* standard set by IBM, the family also includes suppliers such as Compaq, Zenith, Epson and others which produce excellent systems.

[21] Text files on all MS/DOS systems, including those made by IBM, uses a superset of the ANSI code for information interchange, commonly referred to as ASCII.

[22] The use of Lotus 1-2-3 appears to be pervasive in statistical offices in developing countries. As a result, its specialized WKS format is a good candidate for the international transmission of statistical tables, surveys, and other statistical data structures as long as their size is within approximately a 360 KB limit.

files and Lotus should be considered as an interim de facto standard[23] **for the international transmission and interchange of statistical data collections of moderate size**. While such a standard is not all inclusive -- there are many data collections for which it is not adequate -- it has the merit of satisfying a large class of interchange activities well, and it can be supported in almost all if not all of the national statistical offices in the world. The international statistical community should take advantage of its de facto adoption, and build additional layers of standard on top of it where necessary to support more complex information transfer requirements.[24]

The use of common recognized working standards for data transmission at the physical medium and recording format layers is a necessary but not sufficient condition for effective and widespread machine readable data dissemination. Within countries, the effectiveness of disseminating data in such a form depends upon having data users who are also computer users and who can exploit machine readable data directly and without conversion, transition, or learning costs. Feedback from such users is important to minimize the time and cost required for effective exploitation of the data. In such cases, producer and consumer have a strong joint interest in the identification of computer tools that accomplish this objective.

New responsibilities assumed by microcomputer users. The introduction of microcomputer technology in an organization implies a significant transfer of responsibility from a central computer support staff to individual users of the microcomputer systems. The transfer of responsibility has both benefits and costs for the user and for the organization.

In organizations that use mainframe and minicomputer systems, central computer support staffs provide a variety of functions that may be invisible to the users of the system. These functions include:

---

[23] It is not now clear how long an interim standard this might be. The 5 1/4" diskette drive and the MS/DOS formatting standard are likely to survive for the next several years, which should be quite useful for catalyzing statistical data transfers internationally. Such a standard could evolve, following IBM's evolution of its system units to the PS/2 and a 3 1/2" form factor diskette, or it could be replaced with a competing standard. For the moment, however, it is probably the most accepted *de facto* standard in developing countries and as such should be exploited fully.

[24] This recommendation is offered with some concern that, while it may serve to improve the manner in which important statistical data processing activities are executed, it may also serve in the future to block progress to a more productive computing environment. Organizations employing computing equipment often exhibit considerable inertia in taking advantage of newer developments. Thus, past computing technology that was appropriate at the time of its introduction is retained well beyond its appropriate lifetime, even taking into account the sometimes sizeable transition costs in moving to a different model of processing. For example, the significant introduction of microcomputing technology could have occurred 3-4 years sooner in many developing countries than it actually did; this diffusion of technology was delayed by, *inter alia*, lack of information and experience, anxiety regarding such a substantial shift in processing models, and an inertial type of comfort in remaining with tried and true standards that served the organization well for a period of time in the past. This behavior is observed equally in developed countries; developing countries, however, have more to gain by identifying and understanding the phenomenon and by doing so, ensuring that they do not become locked into standards and methods of data processing for longer than those methods and standards serve the country's interests. Thus, the MS-DOS recommendation made should not block initial investments in alternative technologies, because these newer technologies will certainly evolve to render it relatively ineffective.

1. performing a requirements analysis for the organization as a whole and specifying the characteristics of the system to be acquired;[25]

2. negotiating with acceptable suppliers and evaluating their responses to the request for proposals to supply hardware and software components;

3. designing and constructing or adapting the physical site for the computer system, including electrical power requirements and possibly backup power generation, and air conditioning and other environmental requirements;

4. co-ordinating and assisting in delivery, installation, and acceptance testing of the system;

5. providing system programming skills and services, generating the operating system and updating it periodically, and installing application program packages;

6. providing system resource management, including allocations for scarce resources such as disk space;

7. providing user consultation and education of various kinds, and possibly a more turnkey operation including system analysis, design, programming, testing, implementation and documentation;

8. providing operating services, including machine operation, tape and disk volume storage and control, file backup, and an inventory of expendable supplies;[26]

9. providing fault diagnosis and hardware maintenance services (either internally or through a external supplier); and

10. monitoring the market for additional or replacement hardware and software to maintain a cost effective data processing operation.

Users of microcomputers are faced with the same set of responsibilities, although at a much lower level, and often with substantially less risk to themselves or to their organization if non-optimal decisions are made. Nevertheless, these responsibilities have been by and large transferred to users albeit on a considerably smaller scale, and users must be aware of them if decentralized microcomputer use is to be viable within an organization. This issue becomes immediately visible when malfunctions occur, and it is up to the user to diagnose the problem and solve it. Organizations with non-programmer users may want to retain a central data processing services staff for dealing with the more unusual occurrences, while leaving the operational responsibility for substantively oriented work with the user.

---

[25] This task is considerably simpler for the microcomputer user. Not only are the tasks simpler and more homogeneous, but given the relatively low level of expenditure, often several different small computers can be purchased. Such diversity of hardware or software may be more effective in meeting institutional needs than earlier purchases of single systems.

[26] The importance of device and file backup, routinely performed centrally at most mainframe installations, is routinely ignored by most microcomputer operators, and at their own peril. Since the data processed on such machines is institutional, distributed data base management must ensure that microcomputer users are not only conscious of the possibility of data loss through failure to backup, but also perform backups on schedule

Issues of data control and ownership. The data capital of a national statistical organization is one of its most important assets. The data are collected, stored, and used on behalf of and in support of national goals. Although the data are generally placed in the public domain to obtain full utilization from them, the statistical organization has the responsibility for their integrity and accuracy. Responsibility for the data collection is organizational, even though specific persons and groups in the office have the responsibility for collecting and processing them.

Centralized computing environments assist in meeting the above responsibilities. When the data are processed on one computer and stored there, they can be managed and kept secure by the computer operations staff. Likewise, central processing procedures help to ensure both formally and informally that the data maintain coherence and that there are identifiable versions of data, including the version that is officially released. If central data base management systems are used, they further encourage consistency of the data capital through their data base management function and through system wide data dictionaries which are used by programs addressing the data base.

The content of organizational data bases within decentralized computing environments can be more difficult to control, especially when decentralization extends to collections of individual microcomputers. The essence of the typical microcomputer environment is that it is *not* shared, that it *is* individual, and that it supports directly the individual user's productivity rather that organizational objectives. Within such an environment, it is easy for copies of data sets to proliferate; the copies may be of the same version or they may differ, with varying results for processing steps that involve them.[27]

The challenge of obtaining true organizational productivity from such an environment -- in addition to individual productivity -- is to reconcile the individual's personal interest in the data that he or she uses with a framework that supports the organization's interest in the integrity of its data. This is a complex issue, and it will not be solved easily. Recognition of the importance of the issue is the first step in addressing the problem, but further management controls are probably required. Balancing these interests for the overall benefit of the statistical service must be done individually in each country depending upon its circumstances and modalities of operation.

Software as Intellectual Property. Within the last several years, microcomputer systems have become important to many branches of government, both in developed and developing countries, as a source of efficient and productive computation in many areas. Rapid technical progress in the underlying technologies have made such systems affordable not only to government agencies and commercial firms, but also to individuals.

The rapid expansion of the market for microcomputers has made possible a mass market for high quality software which runs on those systems, and software producers

---

[27] As an example, consider what could happen in a decentralized environment in taking a survey of industrial production. Suppose the master register of producers were updated in one place, and copied several times to support separate groups drawing the sample, processing the results and following up non-responses. If the separate registers were not kept in synchronization, then bias might be introduced into the sample, errors might be introduced in the processing stage, and non-respondents for follow up might either not have been selected or they might no longer exist. The result could be either undetectable error or administrative confusion, or both.

have responded by offering a wide variety of products at very low prices relative to the value they provide their users. Without such software, microcomputers would be useless. The cost of this software is a legitimate and essential cost that must be borne in order to provide a productive environment for the use of microcomputer systems.

While software prices may be regarded as reasonable within the context of a developed economy, developing countries and/or international aid agencies may regard the same prices as exorbitant or unaffordable within the context of limited development aid or operating budgets. This conflict between wanting to use software and not being able to afford it may lead to the temptation to make copies of software already purchased rather than purchasing new copies.

Computer software is an important form of intellectual property. It provides users considerable value at low cost. Unfortunately, like a book, it can be copied quickly and easily. Doing so with the intent to use the duplicate copy on yet another computer will in almost all cases violate the license agreement entered into by the original user and will almost certainly be in violation of copyright laws.

Through their activities, statistical offices produce significant intellectual property. The information summarized and prepared, often at substantial cost and with substantial effort, is of great use in preparing development plans, formulating policies, and in assessing progress toward meeting national goals. As creators of information, statisticians are acutely aware of the cost of producing information and the value that it provides.

Statistical information produced by national statistical offices is generally considered a public good; its production and distribution are supported by the government, and it is generally placed in the public domain so that full and free use can be made of it. Private firms developing software have no such subsidy, and depend upon license fees to maintain their development and support activities. Developing countries depend heavily upon microcomputer software products to improve the productivity of, *inter alia*, their national statistical offices. It will be important for software producers to perceive that public institutions in developing countries respect others' rights to intellectual property. Illegal copying of software is a significant problem in developed countries, and the consequences of users in developing countries being perceived as a part of this problem could be serious.

Software producers are aware of the difficulty that some customers have in paying for software, and have often made price concessions for either volume procurement or for procurement by students, staff and faculty in educational establishments. Software producers are generally not unsympathetic to the problems faced by developing countries. When such software is important, the companies can and should be approached either directly or through appropriate bilateral or multilateral aid channels to negotiate more favorable and affordable prices.

Revised selection criteria for computer processing environments. The degree to which microcomputer technology is applicable to national statistical office operations will change over time. In general, the technologies underlying computing hardware and the expanding market for mass distributed software will change faster than a statistical office's data collection, processing and dissemination plans. One would

therefore expect that a greater fraction of the office's processing will be tractable within a microcomputing environment as time progresses.[28]

Advanced statistical offices as well as offices in large developing countries will still require mainframe or minicomputer resources for some time to come. Microcomputers still lack the very large file storage and sequential file processing capability that larger computer systems support well. This relative advantage of larger computers will diminish over time for applications in statistical data processing.[29] The difficult job of the statistical office is to match activities involving statistical data processing with the appropriate computing environment to perform the activity efficiently and effectively. Such decisions over time are not trivial ones; furthermore, they are interlinked with the office's investment strategy in computing equipment which is quite important given the limited funds most developing countries have to invest in such facilities.

VI. STATISTICAL DATA BASES AT THE NATIONAL AND INTERNATIONAL LEVEL

The development of statistical data bases is important for developing countries. Although the collection, analysis and publication of statistical data on a piecemeal sectoral or cross-section basis can yield important information, the data are capable of much more useful exploitation than that. The traditional model of a statistical office as a collector and publisher of information -- as opposed to the notion of an information service specializing in the collection, protection and dissemination of data capital -- is in large part responsible for blocking growth into the area of data base development.

Although the set of publications produced by a national statistical service is an early form of data base, albeit a printed form, the term *data base* as commonly understood now generally implies data stored in machine readable form and generally computer resident. The notion of data base is a general one in computing technology, and refers to an integrated set of data relationships, often complex, that are addressed indirectly through another set of data relationships, generally called a data dictionary.[30] The term data bank is

---

[28] This tendency has been observed by the United Nations Statistical Office between countries over the last 10 years with respect to technological approaches to assistance in population census data processing. Prior to 1979, no census projects used microcomputers for processing. In 1979, microcomputer based processing was initiated in the Cape Verde Islands, which had a population of about 400,000 persons. Since then microcomputers have been used for processing in, inter alia, Laos and Burkina Faso, and at present plans are being made to support total microcomputer based population census processing in 1989 in Kenya, which has an estimated population of 24,000,000. This trend is expected to continue.

[29] One can already observe in developed countries the encroachment of microcomputer oriented large data storage devices into what has been mainframe territory for years. In particular, optical storage devices are now being used to store and process files of the order of 100-500 MB, a size that would have taken 3-12 reels of magnetic tape several years ago. The present competition is between CD-ROM disks, WORM (Write Once, Read Many times) disks, erasable optical disks, and digital audio tape media. One or more of these media will develop to provide effective large file storage for microcomputers, thus eroding another computer capability that was previously only associated with mainframes. The challenge for developing countries is to be able to understand when such technologies become sufficiently mature to incorporate into their computing environment, and then what specific activities and problems to apply them to.

[30] Statistical data processing literature often refers to the contents of the data dictionary as metadata, that is, data describing data.

an earlier term whose origin is more directly related to statistical data processing. The term is a vague one, referring to a significant or comprehensive collection of data in a specific subject area such as statistics, and has no implications regarding the structure of the data collection.[31] The notion of data bank has a somewhat more passive connotation than the notion of data base; yet another concept that is roughly synonymous, that of *information system*, connotes an active process orientation rather than an orientation toward storage and preservation. All of these terms are broad and cover a heterogeneous set of realizations of the concepts.

The terms *statistical* data base, bank, and information system encompass the notion of electronic manipulation and dissemination of subsets of the data collection. The existence of flexible computer based statistical data bases allows views of the data that are not related to the original purpose of the data collection. The addition of data collected from other sources, say from other countries, can provide additional and comparative data for a statistical office. Manipulation of the data does not have to be confined to numeric form or to tabular presentation. Effective tools allow two and three dimensional graphs to be created and manipulated; data collections can be explored in the form of related collections of graphs. The geographical side of data can be exploited; much data collected by statistical offices has a geographic attribute and can therefore be located on a map and associated with other data measured at that and other locations.

Statistical data bases provide, *inter alia*, another form of data dissemination, and one that is more flexible than the classical statistical publication. Statistical data banks should be able to provide statistical outputs of various kinds when they are needed, using the most current data available. Further, in contrast to computer programs that produce specific publications having fixed formats, statistical data banks should respond effectively to *ad hoc* requests for tabulations and other outputs.

The effectiveness of a national statistical data bank depends not only upon the generality of the storage system and the flexibility of the retrieval system, but also upon the ability and willingness of consumers to use the system. Aspects such as consumer education, physical access to the system, and turnaround time are important in promoting consumer acceptance.[32]

The present agglomeration of different and incompatible data bank systems at the regional and agency levels internationally interferes with the objective of distributing

---

[31] In some countries having a significant concern with issues of individual privacy, the term has developed to have a somewhat pejorative connotation at times in the past because of an implication of loss of privacy through state matching and analysis of individual data.

[32] The issue of assessing consumer demand for statistical products is relevant in this context; unfortunately it is sufficiently broad in scope that it could be the subject of an entire paper. An assessment of this demand is important if a national statistical service is to allocate scarce resources among both ongoing activities and new developments such as computer based statistical services. To the extent that the national statistical service plays a direct integrated role in policy planning and assessment within the government, both the type and level of user demand will be easier to ascertain, since the statistical service will be called upon to address issues relevant to current policy issues. Assessing demand at the level of publications is considerably more difficult, and does not address the latent demand for other forms of statistical products that might be more useful, such as machine readable products, products available through computer based statistical systems of various kinds, and more integrated statistical systems such as the geographic information systems discussed in the next section.

statistical data in a timely and effective manner. Standards for disseminating data within the international community need to be reviewed to identify a small number of data structures that can transport most of the data collected by countries and reported to international agencies. At the physical level, the de facto existence of MS-DOS microcomputers in almost all statistical offices should be recognized, and statistical interchanges standards should at this time, where possible, be built upon that standard.[33] For representing information, ASCII files would be appropriate; alternatively, spreadsheet files based upon Lotus 1-2-3 are probably almost as acceptable, since the Lotus 1-2-3 program appears to be nearly as ubiquitous as MS-DOS based systems. For data entities that are at present too large to fit into MS-DOS diskette files, additional standards will be needed.[34] However, it is generally possible to divide a collection of information into subsets which will fit into an MS-DOS diskette file.

The statistical data banks of the specialized agencies and regional commissions are built upon a variety of mainframe platforms, which adds somewhat to the problems of importing and utilizing the data. Some data banks require special software which is only available for computers from a specific supplier or is otherwise idiosyncratic. In addition to the importance of the data transfer format being widely accepted, there would be some advantage in having the various statistical data bank systems resident on the same or similar computer hardware so that the software for storage, retrieval, analysis and publication of the data could be moved from machine to machine and therefore from country to country. This goal was never even approached when most computers in statistical offices were mainframes, but it may be possible to achieve it at least partially now because of the dominance of MS-DOS machines. Within several years, these machines will have large storage peripherals, and MS-DOS will either still be standard or will have evolved into OS/2 or some other operating system standard. The size and power of the resulting microcomputer configurations will be sufficient to support a significant statistical data bank suite of programs of considerable complexity, and this may be sufficient to entice private and/or public investment in such a software system.

Geographic Information Systems. The potential for graphical representation of data has always existed, but the sets of tools that have historically been used to process

---

[33] The use of an MS-DOS microcomputer standard for transmitting information in no way implies that the data bases at either end must be limited to MS-DOS systems, although in the less developed countries that is likely to be true. Statistical data bases in developed countries and international agencies generally use sophisticated data management software executing on larger computer systems, which is appropriate for their work. However, the use of such systems should not add to the difficulty of either collecting data from or disseminating data to developing countries. The nature and sophistication of the large data management system should be visible to the country only to the extent that the country can benefit from knowing more about it.

[34] In principle, standards for transmission and interchange of statistical data are independent of any particular technology, and could have thrived equally well when mainframe technology was dominant. The reasons for lack of such standards are many, including legitimate physical differences at the lowest level among equipment produced by various manufacturers. Individual standards were generated by multiple organizations in both public and private sectors based upon their ability to satisfy particular objectives. The slow growth of interest in dissemination of data in any form beyond the printed page caused machine readable data dissemination activities to be assigned a low priority in many organizations; the standard for dissemination was often the standard used for internal statistical production, and it was exported without careful consideration of whether it was effective in meeting the recipients' needs.

statistical data make it difficult to obtain useful graphical output.[35] Some microcomputer based software products can now help to reverse that trend. Spreadsheet programs such as Lotus 1-2-3 and Excel have graphics capabilities that can be used to advantage in displaying statistical data. ASCII files can be used to transfer data between data base storage programs and a variety of stand alone graphics display programs. In addition, programs such as Data Desk II are emerging which provide the capability to view and manipulate data powerfully in two, three, and four dimensions.[36] The use of statistical data capital as part of a geographic information system must not be ignored, and relates to the eventual use of the statistical data produced. While the classic form of statistical output consists of publications, tables are poor ways to present data which has geographic attributes and which can be visualized as data defined over regions and areas in the country.

Geographic information systems refer to computer based systems that bring together information with a spatial component and allow the information to be manipulated and displayed in thematic map format. It allows information of different sorts to be merged, new attributes computed, and thematic maps made on the basis of values of one or more spatial attributes. The emphasis upon explicit graphical display of statistical data in a thematic map format makes possible a new and powerful dimension of information representation that can be considerably more productive for many users of the data.

A simple example may be helpful. Consider a map of a country having the boundaries of all administrative subdivisions present down to the county level. Such a map might contain between 500 and 2,000 subdivisions. From the population census data base, one can extract population summary tables for each of these areas. From other data bases developed by the country, one could determine the number and capacity of schools at various educational levels. By making some assumptions about the demand for education if the supply were there, one could compute local demand functions for continuing education, and display them as a function of their deviation from the norm. The result is a map whose thematic information is derived from separate statistical data collections, brought together and displayed *de novo* for a special analysis.

---

[35] Early mainframe technology was oriented strongly to the processing of numeric data, with concern for processing of text growing only relatively recently. Both forms of processing were essentially character-oriented, and both the hardware developments and the software tools associated with the technology reflect these concerns. The MS/DOS based microcomputer family essentially continues this tradition, with screens oriented to displays of characters and operating systems built upon complex character strings. Computing paradigms oriented around processing and display of visual images have existed for some time, but have been quite expensive on a relative basis, and prices for corresponding software have been high because of greater complexity of production and more limited markets. This situation has changed radically in the last several years with the strong emergence of other microcomputer architecture paradigms that make extensive use of graphics concepts (examples include Apple, Atari, Commodore, and X-Windows under Unix). At present, there is little evidence that this alternative paradigm is either known or understood by many statistical offices.

[36] Data Desk II is one of a class of programs implemented on the Macintosh rather than MS-DOS machines because of the more powerful user interface and graphics capabilities of the Macintosh. Just as the MS-DOS environment will probably become an accepted standard for the more classical forms of statistical data processing on microcomputers, the Macintosh is likely to and deserves to serve a very powerful adjunct presence in statistical offices to perform processing that is at present somewhat non-standard, but is so powerful and important that it should not be neglected.

It takes little imagination to extend the example further. Fields of statistics dealing with the production and distribution of goods might benefit from an understanding of how the transportation facilities of a country are positioned to assist in distribution. Maps involving distribution could easily be overlaid with physical infrastructure data describing transport networks: highway and roads, railroad lines, canals and waterways, airports, and other means of transport. Fields of statistics dealing with balances, such as food and energy, can use maps to advantage by displaying thematically the geographic distribution of production and consumption. Adding the paths and capacities of distribution mechanisms displays graphically economic relationships of importance to the country.[37]

The generation of even relatively complex thematic maps using large configurations of existing microcomputers is now possible and becoming easier and less costly as software tools increase and as the underlying hardware costs decrease. Powerful programs for supporting aspects of geographic information systems are now available on large MS-DOS environments. The thematic map is for many purposes a superior data integration and display tool for users of statistics than are standard sectoral tabulations, and such maps can be generated relatively easily with existing software bases.

Geographic information systems can be used to prepare material for official publication. Such images, when designed appropriately, can enhance both the readability of a publication and assist the reader in understanding the statistical relationships presented. Many official government publications already use maps to advantage now.

Computer based geographic information systems can also be used to prepare *ad hoc* maps to assist planners and policy makers. This capability may be more important in that, in conjunction with a statistical data base, it can be used when and where needed to study problems of specific interest. It places the statistical service in a proactive rather than a static role, as a partner with other governmental agencies in satisfying the need for statistics in rapid response to areas of concern.[38]

The statistical service itself requires a cartographic data base for population census enumeration as well as sample frame selection. In some countries, a central cartographic agency has primary responsibility for mapping and making derivative maps available to the statistical authorities. However, in many developing countries, there is no such authority and the statistical agency must accept the responsibility. Since the time required to digitize a map for a geographic information system is approximately equal to the time to draw it,

---

[37] These examples describe maps that combine statistical data and administrative records. Other sources of data can also be integrated into the map model: remotely sensed data collected by satellites can be used to determine land use patterns, for example. Population subsets could be used as thematic mapping attributes for studying planning and policy issues relating to that group. Examples are the provision of health facilities and services compared to incidence of diseases, and the geographic distribution of individuals in various occupations, such as teachers, doctors, and technical specialists.

[38] The ability of the statistical service to assume a proactive role by responding rapidly to requests is independent of the existence of an information system specifically incorporating a geographic component. However, many national concerns have a geographic component, and inattention to geography lessen the effectiveness of the statistical service to respond. Furthermore, the map framework is sufficiently general to be able to integrate information from other sources as well as provide pictures for assimilation instead of tables. If the pictures can be produced rapidly at low cost, then interactive dialogues can take place between planners and the information system, allowing a powerful mode of investigation and exploitation of the statistical data base.

statistical offices should exploit the potential of geographic information systems and graphics capabilities on microcomputers to study shifting their cartographic work and data at least in part to microcomputers.[39]

At the international level, data bases provide the same feedback to international planners and aid agencies regarding the state of the world and the effect of international programs that national data bases provide to national planners. International statistics also provide measures of the similarities and differences between countries, and contain a statistical base for analysis of the differences. International data bases provide an explicit geographic component -- the individual country.

Organizations managing statistical data bases at the international level can assist countries by providing them with both national and comparative analyses that might be impossible for the country to execute independently. If parts of the data collection could be transmitted to the country itself, then the country could be empowered to conduct its own comparative analyses. Such transmission and use is possible if there is an agreed upon data transmission format and if the country has the computing capability to read the data in its machine readable form and manipulate it in the manner desired.

Organizations with computer based data bases can assist countries in the process of reporting their data to international agencies. Individualized questionnaires can be printed for each country, containing data for past time periods; this gives the country an opportunity to update previous data as well as providing an approximate comparison for the new figures being reported. A few international organizations have taken the next logical step in a microcomputer environment by distributing questionnaires in spreadsheet form, with both previously reported data and mechanisms that provide both approximate and exact consistency checking of the data being entered. To the extent that this is successful, it detects possible reporting errors at the point of entry rather than at the time of receipt by the collecting organization, which should increase the quality of the data collected. Such steps reduce the burden upon respondents who report data, and allow them to use the resources saved in productive internal activities.

VII. CONCLUSION

Progress in statistical data processing is an important ingredient for effective planning and development in developing countries. It depends upon appropriate and effective exploitation of new computing technology from which everyone benefits on a continuing basis. Such progress occurs and is modulated by a variety of factors specific to individual countries, including infrastructure considerations, incentives for learning and working, and the nature and extent of relationships among organizations having a stake in information technology in different sectors of the country.

The specific path of development of statistical data processing in a country is to some extent dependent upon the particular circumstances and needs of the country. Nevertheless, at the present time there are a number of insights and strategies for using information technology to improve the effectiveness of national statistical services that appear to have higher than average expected payoff for many developing countries. Such

[39] The emphasis upon MS-DOS as the standard microcomputer operating system has slowed a good understanding and appreciation of these possibilities. Computers having flexible, standard and adequate graphics capabilities, such as the Macintosh, have a decided advantage for this type of work.

insights and strategies are worthy of examination by individual countries to determine the extent to which they are applicable to their specific circumstances.

In addition to the various suggestions presented in the previous sections of this paper, the following recommendations are offered for evaluation:

1. Microcomputer based information technology now offers substantial hardware and software processing capabilities which are directly applicable to tasks in statistical data processing. It is the most important emerging computing technology today for intensive exploitation by statistical services in developing and developed countries. Developments in optical digital storage technology are also quite important, and are emerging as products within microcomputing environments that will be extremely important in the near future for national statistical services.

   While many developing countries are aware of the potential of microcomputing technology and have started to exploit it, insufficient information is available to countries describing in detain how this technology is being used in other developing and developed countries. Articles, books, case studies, documentation, programs and short courses would all be helpful if distributed both to and within developing countries. Training activities financed by technical co-operation programs oriented toward toward co-operation among developing programs could also be exploited to disseminate available knowledge more widely.

2. Developments in both hardware and software have improved substantially the environment and computer based tools available for processing statistical data. Statistical services should give serious consideration to beginning to shift the processing responsibility for a significant amount of processing back to the subject matter specialist, redefining the role of the data processing service to be increasingly one of providing specialized assistance, information, and advice.

3. Ameliorating information poverty in statistical services and their supporting computing services in developing countries is critical for improving statistical data processing. A mixed strategy is necessary to accomplish this. The benefit-cost ratio of an improvement in supplying relevant information and in assisting national professionals to understand how to exploit it is relatively very high among forms of technical assistance.

   There is no substitute for organized and focused distribution of information to those who require it and can exploit it. The notion of a statistical data processing newsletter still appears to be a good one, assuming that it can be administered informally and efficiently, and that useful information can be disseminated in a timely manner. The technical means for doing this seem to be available; it is less clear how an effective administrative mechanism would be chosen or created and how it might operate.

   Developed country missions and the offices of both national and international aid agencies in developing countries could assist with the general problem of ameliorating information poverty, since in some countries they contain a significant part of the data processing expertise resident in the country. Libraries and cultural centers established by developed countries provide a similar type of resource. Missions and aid agencies often have data processing expertise on their staffs; they also have access to a flow of data processing information from their home

organizations, as well as reliable delivery of such information through diplomatic pouch services. The information resources and acquisitions of these organizations as well as the computing skills and enthusiasm of members of their staffs should be utilized in as widespread a manner within the country as possible .

4. Issues of data integrity, confidentiality, ownership and control assume new dimensions in a decentralized environment, and must be understood and satisfactorily dealt with by the national statistical service in order to support sustainable progress in the exploitation of new technology. More in-depth assessments of this issue, including case studies, would be very useful in providing guidance to countries in establishing and managing decentralized data processing operations.

5. The international statistical community should recognize the near universal use in statistical services of new low level de facto standards for data interchange and exploit them to improve the efficiency of international data collection, interchange, and distribution activities. International agencies are well positioned to take a lead role in discussing this issue and proposing standards, both informal and formal, short run and long run, in this area.

6. Gains realized from introducing specific uses of the current generation of microcomputer technology should not be allowed to obscure the more sizeable future benefits available from taking a considerably broader view of microcomputing technology. Modern information technology provides inexpensive and effective means of data dissemination and presentation that transcend traditional methods of data dissemination through publication. In particular, models of processing dealing with image manipulation in a variety of contexts have enormous potential for making statistics usable in a dynamic manner to key users of official statistics.

   Developing countries need exposure to these possibilities for dissemination and presentation now so that they can plan to exploit them effectively on a timely basis within the context of their own needs. National and international aid agencies are well positioned to propose projects and components of projects that might provide good case study material in this area. Countries could benefit significantly from knowing more about developments that have occurred in this area in both developing and developed countries and in both the private and the public sector.

7. The traditional method of dissemination of statistics through publication can now be integrated much more closely with the statistical production process inexpensively, with considerable potential savings of time and money and with no decrease in quality. Microcomputer based publishing tools should be integrated into the statistical production process. Countries should be encouraged to study the applicability of this technology in their own special circumstances; aid agencies should consider funding and supporting such activities in countries where there is interest and it appears that the infrastructure can support it.

8. Training strategies in national statistical services can benefit from reexamination within the context of the above strategies. Training for subject matter specialists should now include intensive use of common computer productivity and statistical data processing tools for less complex statistical procedures. Government data processing specialists should have their roles broadened and deepened to reflect the increasing level of professionalism necessary to support a wide range of applications

with more substantial depth of experience. Finally, explicit support for an intensified level of training for both groups is required to support the restructuring of responsibilities involved and to realize effectively the promise of information technology.

The issues raised in this paper are important ones, but they by no means exhaust the set of issues on which meaningful discussion and investigation could take place in the field of statistical data processing. Moreover, many of the issues raised have been discussed at a more superficial level than is desirable because of limitations of space and time in preparing this paper. Nevertheless, they are presented as a beginning, with the hope that some of the ideas presented and discussed will make a meaningful contribution to helping developing achieve their important national goals and priorities.

VIII. SELECTED REFERENCES

Bennett, John M. and Kalman, Robert E., eds. *Computers in Developing Nations*. Proceedings of the IFIP TC-9 International Seminar, Melbourne, 13 October 1980. North Holland Publishing Company, Amsterdam, 1981.

Carlson, Beverley A. and Sadowsky, George, "Survey Data Processing in Developing Countries: Selected Problems and Issues," *Proceedings of the 100th Session of the International Statistical Institute,* Amsterdam, August 1985.

Cotton, Paul, "A Comparison of Software for Editing Survey and Census Data," Presented at *Symposium 88: The Impact of High Technology on Survey Taking*. Ottawa, Canada, October 1988.

de Clercq, Robert, "Traitement des Donnees d'un Recensement par Micro-ordinateur," United Nations Statistical Office, Project INT/88/P09, June 1988, 45 pp.

Doucet, J. E., "The Use of New Technologies at Statistics Canada," Presented at *Symposium 88: The Impact of High Technology on Survey Taking*. Ottawa, Canada, October 1988.

Economic Commission for Africa, *Report of the Training Workshop on the Use of Software Packages in Statistical Computing for English-Speaking Government Statisticians and Planners: Addis Ababa, 21 September-2 October 1987*. United Nations Economic and Social Council, Document E/ECA/PSD.5/47. 28 December 1987, 20 pp.

Food and Agriculture Organization, *Programme for the 1990 World Census of Agriculture*. FAO Statistical Development Series No. 2, Rome, 1986. 119 pp.

Food and Agriculture Organization, Microcomputer-Based Data Processing: 1990 World Census of Agriculture. FAO Statistical Development Series no. 2a, Rome, 1987. 84 pp.

International Statistical Programs Center, *CENTS: Installation Guide for PC Systems*. U.S. Bureau of the Census, Release 5.0, September 20, 1988..

International Statistical Programs Center, *CONCOR Edit and Imputation Ssytem: Installation Guide*. U.S. Bureau of the Census, Release 3.0, September 20, 1988.

Johns, Robert, *Fundamentals of Computer Processing for Statisticians.* United Nations Economic and Social Commission for Asia and the Pacific, Publication ST/ESCAP 537. August 1987, 70 pp.

Keller, Wouter J. Keller and Bethlehem, Jelke G., "Computer Assisted Survey Processing on PC and PC-LAN," Presented at *Symposium 88: The Impact of High Technology on Survey Taking*. Ottawa, Canada, October 1988.

Ruth, Stephen R. and Mann, Charles K., *Microcomputers in Public Policy: Appplications for Developing Countries*. AAAS Selected Symposia Series, No. 102. Westview Press, Inc., Boulder, 1987.

Sadowsky, George, *The Use of Microcomputers for Census Data Processing*, Statistical Office, Department of International Economic and Social Affairs, The United Nations, New York, 1989.

Sadowsky, George, "Statistical Data Processing in Developing Countries: Infrastructure Considerations," *Computers, Environment and Urban Systems*, (to be published in December 1989).

Statistical Commission, *Special Issues: Communications Between National and International Statistical Agencies and the Development of Data Bases*. United Nations Economic and Social Council, Document E/CN.3/1987/2. 2 December 1986, 22 pp.

Toro, Vivian and Chamberlain, Kathleen, "The Use of Microcomputers for Census Processing in Developing Countries: An Update." International Statistical Prrograms Center, U.S. Bureau of the Census. Presented at *Symposium 88: The Impact of High Technology on Survey Taking*. Ottawa, Canada, October 1988.

United Nations, *The Application of Computer Technology for Development*. Department of Economic and Social Affairs, The United Nations, Document E/4800, ST/ECA/136, 1971. 122 pp.

United Nations, The Application of Computer Technology for Development: Second Report of the Secretary-General. Department of Economic and Social Affairs, The United Nations, Document ST/ECA/176, 1973. 89 pp.

Westlake, Andrew, *The Impact of 'New Technology' on Statistical Offices: Hardware, Software and Systems for Survey Data Processing*. International Statistical Institute Research Centre. August 1987, 79 pp.

UN publications
Some ISI articles, publications
WFS reports
technical background reading
ISSA, Jorge Ortuzar/Westinghouse
CDS-ISIS on microcomputers -- Unesco with IDRC