

ACC SUB-COMMITTEE ON STATISTICAL ACTIVITIES
Twenty-third session
Geneva, 19-23 June 1989
Item 3 of the provisional agenda

SA/1989/2
13 June 1989

PUBLICATION AND DISSEMINATION POLICY, INCLUDING
CO-ORDINATION OF STATISTICAL DATA BASES

Report of the Technical Working Group on Statistical Data Bases

Sixth Meeting, Geneva, 11-12 May 1989

Report prepared by the Technical Working Group on Statistical Data Bases

INTRODUCTION

1. The sixth meeting of the Technical Working Group on Statistical Data Bases was attended by representatives of the United Nations, Economic Commission for Europe (ECE), Economic and Social Commission for Asia and the Pacific (ESCAP), United Nations Centre for Human Settlements (HABITAT), United Nations Conference on Trade and Development (UNCTAD), United Nations Population Fund (UNFPA), International Labour Organization (ILO), Food and Agriculture Organization (FAO), World Health Organization (WHO), International Monetary Fund (IMF), International Telecommunication Union (ITU), Advisory Committee for the Co-ordination of Information Systems (ACCIS), Statistical Office of the European Communities (SOEC) and Inter-American Development Bank (IADB).

2. The main agenda items discussed were:

- (1) Standardization for data transfer:
 - (a) Common coding systems for data exchange: country code, others;
 - (b) Issues on linking statistical data bases - issues on establishing an inter-organizational data base;
 - (c) Use of CD-ROM for data dissemination;
- (2) Current status of statistical data bases
 - (a) Issues on the on-line access to data from other organizations;
 - (b) Commercially available data base management system;
- (3) Use of micro-computers in statistical work
 - (a) Software packages for managing a large number of data bases on micro-computer;
 - (b) Experience in desk-top publishing.

I. STANDARDIZATION FOR DATA TRANSFER

a. Common coding systems for data exchange: country code, others

3. The meeting was informed of the experiences in the use of a number of code sets based on the UN/EDIFACT syntax rules which had been agreed within the ECE Working Party on Facilitation of International Trade Procedures. For example, the country code used in UN/EDIFACT and being introduced in ASYCUDA (Automatic System for Customs Data) was the ISO 3366 two-character alphabetic code; the currency code used was the ISO 4217 three-character alphabetic code which was based on the two-character country code and a third character indicating the first letter of the name of the currency, e.g. FRF for French francs; for ports and other locations, a five-character alphabetic code was used as set out in ECE/FAL Recommendation 16 in which the two first letters gave the country code, followed by three letters representing the name of the location (often the same as in the IATA location identifiers), e.g. FRMAR for Marseilles, France.

4. There was also a view that the coding was not the real problem in the exchange of information but the greatest difficulty lay in the fact that different concepts, definitions and standards were used by the various organizations and even sometimes within the same organization. In respect to common standards, it was mentioned that agencies developed their own systems to meet specific work requirements and that they allowed the standard systems to become outdated.

5. Some agencies reported that they used country codes based on the 3-digit numeric country or area codes "Standard Country or Area Codes for Statistical Use (ST/ESA/STAT/SER.M/49/Rev.2) 1982" issued by the United Nations Statistical Office while others were using either the ISO 2-digit alphabetic codes or the 3-digit alphabetic codes. These ISO alpha-codes were also listed in the same publication from the United Nations Statistical Office. The meeting was informed that the United Nations Statistical Office was considering the revision of the paper taking into account changes in the names and preparation of codes since 1982. } New

6. The grouping of countries or areas used in data exchanges were also mentioned as greatly varying among agencies. Different groupings were being made by geographical characteristics, stage of economic development or type of economic systems, etc. It was suggested and agreed that the United Nations Statistical Office compile a list of country codes, including groupings of countries or areas, currently used by the members of the ACC Sub-Committee for the dissemination of their statistics in machine readable form.

(b) Issues on linking statistical data bases - issues on establishing an inter-organizational data base

7. The meeting was informed that the ACC Sub-Committee on Statistical Activities, at its twenty-second session in 1988, discussed the development of an inter-organizational statistical data base for dissemination and agreed that the development of any large-scale data base of that kind would require a considerable amount of time for development and resources. The Sub-Committee also agreed that a more immediate approach would be to select a limited set of series for inclusion in the proposed inter-organizational statistical data base.

8. The representative of the United Nations Statistical Office informed the Technical Working Group that the Statistical Office was currently computerizing the annual publication of its World Statistics in Brief (Statistical Pocketbook) which contained about 40 basic statistical series starting 1980 for all the Member States of the United Nations. This core statistics series would be stored into an on-line data base at the New York Computing Service which can be accessed through the ADV line linking the United Nations offices in the world. The data base will also include information on the computerized statistical series available from international statistical agencies in machine readable form which were contained in the Inventory of Computerized Statistical Data Bases. The Technical Working Group supported this proposal and suggested that the names and telephone numbers of persons responsible for the statistical data bases in member agencies be included in that data base and updated regularly in order to make the communication between users and data providers as easy as possible.

(c) Use of CD-ROM for data dissemination

9. Several representatives informed the meeting that they had produced or had a plan to produce CD-ROMs which could store a large volume of statistical data on a simple disk. The production of CD-ROMs needs to be done by an outside service bureau, the production cost becomes very low compared with that of a few years ago. The meeting was also informed that the actual cost of production of a CD-ROM depended on whether or not a sophisticated software for retrieving and manipulation is prepared by the service bureau. Without such software, the cost would be in a range between \$1,500 and \$3,500, or an average of \$2,500 for preparing a master disk and \$2.00 to \$3.00 or an average of \$2.50 per additional copy produced. Therefore, if 500 copies of CD-ROM's were produced without a particular software, the production cost would be \$3,750 in total or \$7.50 per copy of CD-ROM. The users of this type of CD-ROM would require a micro-computer-based CD-ROM drive, a drive card and DOS CD-ROM Extension (costing \$600-\$700).

10. A representative mentioned that the use of CD-ROM's for dissemination was a cost effective way of disseminating a large volume of statistical data because of the low cost of media used for distribution and reduced expense associated with distribution of the media (e.g. postages and handling).

II. CURRENT STATUS OF STATISTICAL DATA BASES

a. Issues on the on-line access

11. The meeting was informed that more agencies were providing internal users and external international organizations as well with on-line access to their data bases. There were discussions on the problems of user interfaces or how to best serve users who were not experts in data processing but wanted to obtain information from the data base. A representative reported an experience of the use of the SQL (Structural Query Language) on a micro-computer based data base. A SQL-based package was mentioned as a very useful language providing a facility to generate a series of user interface programmes very easily and quickly. A further exchange of experiences in this area will be made at the next meeting.

b. Use of commercially available data base management systems

12. The meeting noted that a number of organizations were interested in the use of commercially available data base management systems because (a) the in-house development of computer programmes for data base management would be jeopardized when those who worked for programme development left the organization, (b) the development and maintenance of an on-line data base required a significant amount of technical as well as financial input.

13. Several organizations reported that they had already adapted or were seriously considering the use of such a data base management system. For example, IMF was using Model 204 system, while UNCTAD recently adopted the ADABAS system. FAO was reported to have identified the following DBMS as possible candidates: DATACOM/DB, DB2, ADABAS, AXIS and M204. The representative of the Inter-American Development Bank explained their experiences in using DB2 system in the Bank.

14. ECE, on the hand, reported that it had adopted as its statistical data base management system the AXIS (Auxiliary System for Interactive Statistics) system developed at Statistics Sweden, not commercially available. AXIS provides a variety of functions for the storage and maintenance of statistical data organized in different structures (matrices and time series) as well as of system meta data and other descriptive information. ECE is currently preparing an English version of AXIS.

15. A representative of ECE reported the current status of the ECE's Statistical Computing Project which had two primary objectives: (a) joint development of software for a higher efficiency of national statistical services and (b) to be of particular benefit to developing countries. One of the activities relating to statistical data bases was the development of the Base Operator System (BOS) which is a data base management system.

III. USE OF MICRO-COMPUTERS IN STATISTICAL WORK

a. Software packages for managing a large number of data bases on micro-computer

16. As mentioned before, some representatives informed the meeting that they were testing micro-computer-based data base management systems. The relational data base packages, such as dBASE-IV, Informix, XDB, Paradox-3, Oracle, were mentioned as hopeful candidates and, in particular, the SQL-based packages such as XDB, were described as providing an easy-to-use yet powerful capability. Their function as a data base server in a network environment was also touched. These issues would also be discussed at the next meeting of the Technical Working Group.

17. The meeting was informed of the United Nations Statistical Office feasibility study on the use of micro-computer diskettes for collecting national accounts statistics from Member States. This feasibility study was conducted by selecting 27 countries which received the diskettes containing all instructions and definitions, a complete set of pre-filled national accounts tables with data back to 1960. Included in the diskettes was a table-oriented software package which permitted to edit, update and print the national accounts tables and their related footnotes. The Statistical Office would annually expand the use of diskettes until all interested countries were included.

b. Experiences in desk-top publishing

18. Many representatives reported that they were using or experimenting the use of the desk-top publishing technology in statistical work. However, it seemed that the applications of the desk-top publishing were limited to the printing of ad-hoc dissemination of a small number of statistical tables or textual materials of limited volume. Because of the software limitation a volume printing of statistical tables was not possible at this time. It was suggested that the Technical Working Group should also discuss the use of laser printing technology capable of higher density output which is a step ahead of the currently popular laser printer on micro computer.

PROPOSED AGENDA FOR THE 1990 MEETING

19. The agenda for the next meeting would cover the same broad categories of agenda items as those discussed at the current meeting but a focus would be placed on the following:

(1) Current status of statistical data bases

- New developments for member organizations
- Issues on the on-line access
- Data base management systems

(2) Standardization for data transfer

- Common coding systems for data transfer
- Inter-organizational statistical data base

(3) Applications of new technology

- Data base management systems on micro-computers or local network
- Use of CD-ROM for data dissemination
- Use of advanced laser printer systems or other electronic typesetting technology for statistical publishing.
- Data base management systems on mini-computers
- Use of Scanner and OCR

TECHNICAL WORKING GROUP ON STATISTICAL DATA BASES

ACC SUB-COMMITTEE ON STATISTICAL ACTIVITIES

11-12 May 1989, Geneva

LIST OF PARTICIPANTS

UNITED NATIONS STATISTICAL OFFICE

Mr. T. Shigematsu, Chief, Statistical Services Branch

ECONOMIC COMMISSION FOR EUROPE

Mr. R. Gentile, Deputy Director, Statistical Division

Mr. A. Galati, Chief, Information Management Systems Service, of OES

Mrs. J. Meliskova, Chief, Statistical Computing Unit, Statistical Division

Mr. D. Bachmann, Data Base Administrator, DEAP

ECONOMIC AND SOCIAL COMMISSION FOR ASIA AND THE PACIFIC

Mr. T. Miyake, Systems Analyst, Statistics Division

UNITED NATIONS CENTRE FOR HUMAN SETTLEMENTS

Mr. G. Karl, Economist-Statistician, Research and Development division

UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT

Mr. Zandee, Chief, Trade Control Measures Information System

UNITED NATIONS POPULATION FUND

Mr. H. Kimata, Chief, Planning and Statistics Unit

INTERNATIONAL LABOUR ORGANISATION

Mr. P. Cornu, Statistical Studies and Data Base Section

Mr. D. Viry, Statistical Studies and Data Base Section

FOOD AND AGRICULTURE ORGANISATION OF THE UNITED NATIONS

Mr. F. Pariboni, Corporate Database Manager, Statistics Division

WORLD HEALTH ORGANISATION

Mr. J. Duppenenthaler, Statistician, Epidemiological and Stat. Methodology Unit

INTERNATIONAL MONETARY FUND

Mr. J.R. McKee, Assistant to the Director, Bureau of Statistics

INTERNATIONAL TELECOMMUNICATION UNION

Mr. M. Minges, Computer Department

ADVISORY COMMITTEE FOR THE CO-ORDINATION OF INFORMATION SYSTEMS

Mr. C. Szpak, Systems Officer

STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES

MR. J.C. Farget

INTER-AMERICAN DEVELOPMENT BANK

Mr. J. Anas, Statistical and Quantitative Analysis Section