

(c) we want Country nomenclature for data use that reflects a change in the format ^{the basket approach} to include related codes / apps information.

ACC SUB-COMMITTEE ON STATISTICAL ACTIVITIES
Twenty-fourth session
Vienna, 18-22 June 1990
Item 3 of the provisional agenda

SA/1990/4

Report on

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

Report of the Statistical Office of the
United Nations Secretariat

I. HARMONIZATION OF METHODS AND STANDARDS FOR
INTER-ORGANIZATIONAL DATA EXCHANGE

A. Country or area coding and classification

At its seventh meeting, the Technical Working Group on Statistical Data Bases of the Sub-committee noted the importance of the United Nations Standard Country or Area Code for Statistical Use in standardization for data transfer among international organizations and that a revision of the 1982 published version of the code would be very useful. The Working Group discussed the inclusion of various organizations' regional groupings in the classifications shown and the importance of periodic updating.

*was in
the
periodic
update*

In the light of the importance to organizations and users of country coding, geographical groupings and other aggregates and their constituent parts, and the frequency with which changes are occurring and can be expected to continue to occur in country nomenclature and aggregates used for a wide variety of statistical applications, the Sub-committee may wish to consider the following steps to enhance the usefulness, timeliness and dissemination of this coding system:

(a) Issuance on MS-DOS microcomputer diskette in spreadsheet and ASCII formats and on-line availability as well as publication as a United Nations document;

*No. of
publications*

(b) Inclusion of additional aggregates in common use and their constituent parts which may be supplied by members of the Sub-committee under their own responsibility;

(c) Ways of keeping the nomenclature and coding system up-to-date and communicating changes to users.

This is a web document on UNSCO practice. Not set up as a standard format will be an on-line document. If not in the sub-committee work to proceed there are 2 jobs - revision of MS-DOS by UNSCO - separate document on broadening of the data.

B. Methods of data exchange

Apart from existing arrangements for access and exchange of large mainframe data sets through the United Nations telecommunications system, a substantial amount of small-scale data exchange is now occurring through exchange of microcomputer diskettes. Up to now, no regular facility has been established in the telecommunications system, at least at New York, for the communication of MS-DOS files through either United Nations or commercial telecommunications links, for example in common word processing, spreadsheet or ASCII formats. Such arrangements could, in principle, considerably facilitate interorganizational data exchange on a day-to-day basis.

Members of the Sub-committee may wish to comment on their organizations' experience and interest in establishing and using such links for transfer of statistics among organizations and with Governments.

II. DEVELOPMENT OF A CORE UNITED NATIONS SYSTEM DATA BASE OF BASIC SERIES OF GENERAL INTEREST

The proposal to develop a core United Nations system data base of basic series of general interest was discussed at the last two sessions of the Sub-committee and at the last session of the Statistical Commission. This data base is being established by the Statistical Office on an experimental basis using the series included in the United Nations statistical pocketbook, World Statistics in Brief, with some additions as proposed at the last session of the Sub-committee (see SA/1989/4, annex).

As reported to the Sub-committee's Technical Working Group at its last session, most of the series have now been compiled on microcomputer spreadsheets. However, they require considerable review, editing and footnoting before they can be released or converted to a data base system. The thirteenth edition of the pocketbook is now being completed for printing and work on data editing for the data base will thus be resumed shortly.

Some issues which have arisen in this work are the following:

(a) Many of the series, for example concerning population, price indices and national accounts, are retroactively revised on an annual or biennial basis. Hence, it is often necessary in

updating not only to add the latest year but to replace some or all of previous years' data with revised series;

(b) Necessary footnoting is painstaking and time-consuming and spreadsheet or data base software to handle footnotes automatically is inadequate;

(c) Most of the social series, at least for developing countries, are only available on a five- or ten-yearly basis, usually from population and housing censuses, while annual data are widely available and essential to include for many of the economic fields. This implies either (i) splitting the data base into economic and social parts, creating an obstacle to the integrated analysis which is a major objective of the data base, (ii) extensive estimation based on interpolation and extrapolation, or (iii) 80 per cent or more blank cells in the social areas.

The Statistical Office hopes to circulate a preliminary version of the data base in spreadsheet format to agencies for comment early in the fall.

III. DISSEMINATION POLICIES AND NEW TECHNOLOGIES

At its last session the Sub-committee expressed a number of concerns relating to the impact of new technologies on organizations' dissemination policies. These included the following:

(a) Access by developing countries to on-line data bases for collection and dissemination;

(b) Adherence by all members of the Sub-committee to agreed guidelines for secondary dissemination of data by international organizations;

(c) Modifications to the guidelines for secondary dissemination which might be called for in response to new technologies.

In this connection, the Sub-committee considered it would be useful to bring up to date the previously collected information on policies and practices in dissemination of statistics data in the international organizations, including pricing. It requested the Statistical Office to set up an informal ad hoc group to prepare relevant material for the Sub-Committee on the basis of reports from organizations, in particular their experiences with

~~The latter is justified~~ In the on-line or not
"PRICING POLICY OF SOURCE ORGANIZATIONS" needs to be broadened
to include public sector as well as private. Some
other agencies provide free files to the
media. data not to a point including and
+ another organization about the
one of them
to give the
other
file?

new forms of dissemination, for example CD-ROM and on-line, bearing in mind the concerns raised in the Statistical Office's report on this topic (SA/1989/4), such as assurance of adequate substantive support and metadata for users and protection of the source organization's income from sale of the same data.

Up to now it has not been possible for the Statistical Office to set up such a group or to canvass the organizations for information in the appropriate detail. However, several of these points were taken up at the last meeting of the Technical Working Group. The Group noted, for example, that a number of experiments with on-line systems and dissemination on CD-ROM disks are in various stages of implementation and evaluation but that few conclusive, positive results had been achieved as yet. In general, the software and documentation requirements for effective communications and metadata interfacing with users do not yet seem to have been satisfactorily met.

On the question of secondary dissemination, the Technical Working Group expressed the view that the guidelines on secondary dissemination "should be broadened in order to include the supply of the source data to the member governments in addition to international organizations; this being dictated by the growing number of requests for such services" (SA/1990/2, para. 5). The Sub-committee may want to consider this proposal in the light of more general questions of machine-readable dissemination policy discussed below.

Beyond the question of secondary dissemination in machine-readable form to Governments, additional questions are arising concerning (i) compilation and dissemination in machine-readable form by third parties, often private, drawing on both printed and machine-readable materials disseminated by international organizations, and (ii) each Government's right to control international organization dissemination of data provided to the organizations by Governments, particularly international dissemination in machine-readable form and on-line. These questions are becoming more pressing as the costs and technology for large-scale copying and dissemination of data in machine-readable form are now within reach of virtually anyone with \$1000 or so of readily-available hardware and software, making machine-readable copying considerably cheaper and more reliable than photocopying. Given these technologies and their costs, these questions also assume a new order of importance because any kind of public machine-readable dissemination opens the door to redissemination and copying in whole or in part with minimal possibilities of control or supervision.

~~There should be no secondary dissemination of data to the public by international organizations, unless the source organization has given its consent.~~
In the above paragraph, it should be noted that
any data

The Sub-committee may wish to comment on the continued salience of the 1979 guidelines on secondary dissemination of data in machine-readable form (see SA/1989/4 and the annex below) in the light of members' recent experience and plans in this field, and request an informal survey of members' policies and practices by circular letter from the Chairman of the Sub-committee, with a view to circulating the responses among Sub-committee members ~~for informal discussion at the time of the next Statistical Commission and review by the Sub-committee at its next session.~~

*(not necessary at time of Stat Comm
sub all sub allow involvement
Stat Comm
not treat SC time as minutes)*

*not comment there
if it turns out to be
desirable can do so
but not specifically plan it*

Annex

Except from Issues Related to Dissemination Policies (SA/1989/4)

The existing guidelines on secondary dissemination of data in machine-readable form are as follows:

"(b) There should be no secondary dissemination of machine-readable statistics to a customer which is not an international organization where the statistics are in the form originally provided. Such requests should be referred back to the original providing organization;

(c) When the data provided have undergone a transformation, or have been combined with other data, the resultant machine-readable product may be sold to a customer which is not an international organization if the original providing organization(s) agree(s). The agreement should be reached on a product-by-product basis and cover the terms and conditions of the."

The above guidelines for the exchange and dissemination of data in machine-readable form were first agreed upon in 1979 when the provision of data on-line basis was not as active as today. Nevertheless, at its nineteenth session in 1985, the Sub-Committee agreed that the existing guidelines were sufficiently broad and generally applicable and no change in them was required. The general principle of the free exchange of statistical data on an inter-agency basis was considered to apply also to the provision of data on a direct on-line basis so that, to the extent that an organization maintained for its own needs statistical data in a direct on-line data base, free access to those data but not computer processing should be permitted to other organizations (ACC/1985/15, para. 12).