



KENYA NATIONAL BUREAU OF STATISTICS

DATA ACCESS AND DISSEMINATION POLICY

November 2008.

Microdata Release Policy

Preamble

Since late 1980's upto early 2000, the capacity of Kenya's National Statistical System (NSS) to produce timely, high quality and relevant statistical information had deteriorated. Reasons for this included declining financial outlays to produce and disseminate statistics, inadequate professional staff at the senior management levels, and poor management. Until very recently, surveys were conducted but not analyzed, or when analyzed, the results were not released to the public in time. Several statistical products such as the Economic Survey and the Statistical Abstract were produced without release calendars, leading to delays in making information available to the public. Production of other statistical products such as the Statistical Digest, the Annual Trade Statistics, and the Annual Education Statistics ceased. Household surveys and industrial production surveys were conducted sporadically, depending on the availability of donor funding, since Government allocations to the former Central Bureau of Statistics (CBS) were inadequate to finance them.

The country's input/output tables and the social accounting matrix were last produced in 1976, while the system of national accounts used in the compilation of national accounts aggregates had not been fully revised to reflect the guidelines of the United Nations System of National Accounts 1993 (SNA93). Results of the surveys which the CBS managed to conduct were released so late that they were often irrelevant in an ever changing environment. The production of statistics through the routine data systems in the line ministries also declined over time. Moreover, production of statistics by various agencies was not well coordinated, resulting in duplication of data collection efforts. In addition, annual symposiums for data users and producers, which used to be held in the 1980s to ensure that data users and data producers maintained an active dialogue, were not held.

Meanwhile, demand for statistics to benchmark and track implementation progress of national development initiatives and programs substantially increased. The opening up of democratic space in Kenya in the recent years, coupled with increased media freedom and coverage of economic and social issues, raised public awareness of the value of accurate and timely statistical information. Government has also

increasingly come to appreciate the usefulness of good statistics for evidence-based policymaking in guiding major Government policies and in monitoring development programs and the Millennium Developments Goals.

To help rebuild the statistical system, the Government prepared a five-year Strategic Plan for 2003/04–2007/08 for the NSS. One of the key recommendations in the Strategic plan is the development of a Data Access and Dissemination Policy to address the issue of data access and use. The Policy for the Release of Anonymized Microdata Files

Objective: the objective of this policy is to define the nature of the anonymized microdata files that will be released, the intended use of these files, and the conditions under which these files will be released.

Rationale: The KNBS aims to support research and teaching about the economic and social conditions of the Kenyan population. In particular, this policy aims to support the needs of specialized researchers and students by providing anonymized microdata files to be used strictly for research purposes. The output from this type of research forms the basis for decision-making about policies and programs in Kenya.

Policy Statement: The KNBS will release microdata files for use by researchers for scientific research purposes when:

- The Director General is satisfied that all reasonable steps have been taken to prevent the identification of individual respondents
- The release of the data will substantially enhance the analytic value of the data that have been collected
- For all but purely public files, researchers disclose the nature and objectives of their intended research,
- It can be demonstrated that there are no credible alternative sources for these data, and
- The researchers have signed an appropriate undertaking.

Definitions

For the purposes of this policy, **microdata** are defined as files of records pertaining to individual respondent units. Microdata files for dissemination purposes differ from those used within the KNBS in that all direct and indirect identifiers have been removed through various anonymisation procedures.

Anonymisation refers to the process of removing direct and indirect identifiers from the survey file to conceal the characteristics and the identity of individual respondents.

Direct identifiers include such information as names, addresses or other direct personal identifiers which must be removed from all files made available to users.

Indirect identifiers refer to characteristics which are shared with several other respondents, and when combined with other information, can lead to compromising the identity of the respondent.

Indirect identifiers may be removed through the use of one or more of the following techniques:

- **Removing variables:** A variable that is regarded as being too sensitive to be released e.g., ethnicity, religion, HIV status may or may not be relevant for policy analysis and can be removed from the file when not required.
- **Removing records:** Extreme values may be removed entirely from the file and the weighting factor adjusted accordingly. This approach is minimized as removal of a great number of records will distort the data.
- **Global recoding:** This involves the aggregation of the observed survey values in pre-defined classes in such a way that individual responses are not visible. This approach can be applied to continuous or discrete variables and to geographic codes.
- **Top and bottom coding:** For numeric or ordinal variables, the highest and lowest values are often very rare and can reveal the identity of respondents. Top coding involves the creation of 'catch-all' categories such as "age greater

than 75” or “income greater than 6000”. Bottom coding involves creating catch-all categories for small values.

- **Local suppression:** This is used when two variables taken together could lead to identifying a unique person. E.g., a 15 year old widow would likely be a unique situation. But, suppressing either ‘marital status’ or ‘age’ would most likely eliminate this problem. Suppressing marital status may be the best outcome for this particular case.

For the purposes of this policy, **dissemination** refers to the act of making microdata files and supporting **metadata** available for **access and use** with the data.

Types of Microdata files

There are three types of microdata files covered by this policy:

- **Public Use Files (PUFs):** Microdata files that are disseminated for general public use outside KNBS. They have been highly anonymised by removing the names and addresses and by collapsing geographic and respondent characteristic details to ensure that identification of individuals is highly unlikely. These files are made available for downloading from the KNBS site to individuals who identify themselves by name, provide their email addresses and agree to abide by the terms and conditions appropriate for a PUF. See Appendix A
- **Licensed files:** These files require that there be a signed agreement between the KNBS and major users, to permit them to access data files that are less highly anonymized and/or more sensitive than PUFs. For these files, all individual identifiers have been removed and some characteristic detail may be collapsed or removed. Licensing agreements are only entered into with bona fide users working for registered organizations. The primary and secondary researchers must be identified by name and a responsible officer of the organization must co-sign the license agreement. See Appendix B
- **On-site Access files:** The KNBS may, in the fullness of time and based on special needs, permit users to work on-site within the KNBS premises. This

would permit access to sensitive data which could not be adequately anonymised for release outside of the KNBS. All data removed from the KNBS premises must be reviewed by an officer of KNBS to ensure that confidentiality is not breached. Researchers will be asked to outline their research objectives. See Appendix C.

Classes of uses and users

The KNBS recognizes that microdata files are intended for specialized users with advanced quantitative skills. This policy distinguishes among the following classes of users and uses:

- Policymakers and researchers employed by line-ministries and planning departments who are engaged in the development of regional and national strategies and programs, including the monitoring and evaluation of these programs under Kenya's National Integrated Monitoring and Evaluation system (NIMES).
- International agencies carrying out special studies aimed at identifying development and support opportunities and in planning, and assessing effectiveness of public development programs and investment in Kenya .
- Research and academic institutes involved in social and economic research
- Students and professors mainly engaged in educational activities, and
- Other users who are involved in conducting scientific research (to be approved on a case by case basis).

Classes of Anonymized Files

This policy recognizes that different products will have different characteristics and will be of varying use to the users depending on their research objectives. The nature of the files produced will depend on the user needs that the survey/dissemination team have identified prior to producing the file.

Notes on degree of anonymization

There are three main types of anonymized files that may be produced under the terms of this policy. The major differences among these files are the levels of geographic and characteristic detail.

- Files that have less geographic detail and less characteristic detail (i.e., geography has been collapsed and variable detail has been reduced)
- Files that have less geographic detail and more variable detail and
- Files that have more or full geographic detail and more variable detail.

If all the survey variables are available in the file and the geography is detailed, then the greater the risk of identifying a respondent. Accordingly users requiring the maximum level of geographic detail may have to be prepared to work with less characteristic (variable) detail.

Files that have had both the variable detail and the geographic codes reduced are invaluable for students and professors for teaching and learning purposes so long as they are easily available to them. Such files can be made available as PUFs from the KNBS web site (see below).

All files will be reviewed by a Microdata Release Committee prior to release

Notes on Access Methods

The nature of the anonymized file will determine the manner in which it will be disseminated. The KNBS will release its dissemination microdata files on its KDA (Kenyan Data Archive) server. PUFs are made available by having users complete an online form in which they provide their contact information. Having contact information enables KNBS to monitor the popularity of its products and enables contact with the users if there are subsequent problems with the file or updates to the survey.

Files requiring the user to sign a license are made available after the user completes and returns an access form which they can obtain from the KDA site. The KNBS may

choose to provide these users with on-line access from a secure server, secure FTP or CD-ROM. More than one version of the KDA server may be used to serve different classes of users.

Responsibilities

Survey task managers: will be responsible for

- identifying the needs of key stakeholders and ensuring the creation of an anonymized file that meets the needs to the user community to extent that is possible under the Statistics Act.
- preparing an initial screening of the microdata file, identify any potential problem areas to be resolved and preparing a submission to the Microdata Release Committee.

The Microdata Release Committee: will be responsible for :

- reviewing all requests by KNBS survey task managers for the release of anonymized microdata files, using established criteria;
- approving all files for release or providing guidance to the survey manager on how to improve the file before it may be released;
- reviewing requests for access to licensed and on site microdata files by researchers;
- revising the guidelines that managers use to create anonymized microdata files as necessary;
- overseeing the licensing process and resolving issues dealing with possible breaches; And
- making proposals to the Director General on the revision tot his policy as and when deemed necessary.

The Director General: will approve all releases of anonymized microdata files to users based on the advice and recommendation of the Microdata Release Committee.

Sponsors and Line-ministries: will be responsible for identifying their research needs and priorities to the survey task managers prior to the design of the microdata files.

Timing and Release Policy

Recognizing the important needs of users and the importance of timely data, the KNBS will strive to release microdata files within 6-12 months after the first release of data from the survey. Surveys may need to be released in phases to ensure that the survey data have been reviewed and analysed by the KNBS staff prior to their being anonymized for use outside the KNBS. This assures users that the KNBS will be able to answer questions about the survey.

Cost Recovery Policy

It is the policy of the KNBS to encourage broad use of its products by making them affordable for users. Accordingly, KNBS attempts to ensure that the costs of creating anonymized microdata files are built-in to the survey budget. At the same time, KNBS attempts to recover costs associated with the provisions of special services that benefit only a specific group.

The following cost-recovery principles will apply to the dissemination of microdata files.

- Public use files posted in the KNBS data server for broad download will be available without charge.
- Licensed data files will be made available to sponsors without charge providing that the cost of preparing the files is part of the survey budget.
- Line ministries are expected to ensure that the KNBS has sufficient funding to support the preparation of files for their use.
- Users demanding special services are expected to negotiate a cost-recovery agreement with the KNBS

Other Forms of Data

Its important to note that the Bureau also collects, analyses macroeconomic data which includes the national accounts aggregates, interest rates, trade and balance of payments(BOPs), Industrial, labour and earnings, justice etc. All information is found in KNBS reports like the Annual Economic Survey and the Statistics Abstract and the Monthly Economic Indicators reports. All these indicators can be accessed either from the reports or on request.

For Further Information

For further information about access to microdata, please contact:

The Director General,
Kenya National Bureau of Statistics]
P.O Box 30266-00100
Nairobi, Kenya
Fax to254-20-315977
E-mail scanned copy to: www.knbs.go.ke; director@knbs.go.ke

- Appendix A

Terms and conditions of use of public data files

1. The data and other materials provided by the KNBS will not be redistributed or sold to other individuals, institutions, or organizations without the written agreement of the KNBS.
2. The data will be used for statistical and scientific research purposes only. They will be used solely for reporting of aggregated information, and not for investigation of specific individuals or organizations.
3. No attempt will be made to re-identify respondents, and no use will be made of the identity of any person or establishment discovered inadvertently. Any such discovery must be immediately reported to the KNBS.
4. No attempt will be made to produce links among datasets provided by KNBS, or among data from the KNBS and other datasets that could identify individuals or organizations.
5. Any books, articles, conference papers, theses, dissertations, reports, or other publications that employ data obtained from the National Data Archive will cite the source of data in accordance with the Citation Requirement provided with each dataset.
6. An electronic copy of all reports and publications based on the requested data will be sent to the KNBS.
7. The original collector of the data, the KNBS, and the relevant funding agencies bear no responsibility for use of the data or for interpretations or inferences based upon such uses.

Appendix B: Form for access to licensed file from the Kenyan National Bureau of Statistics (KNBS)

Application for Access to a Licensed Dataset

Title and reference number of the dataset(s) you are requesting (use the exact title, year and reference number as listed in our survey catalogue):

Instructions

This form is to be mailed or faxed to the KNBS, with a cover letter printed on the sponsoring agency letterhead.

Mail to: [The Director General,
Kenya National Bureau of Statistics]
P.O Box 30266-00100
Nairobi, Kenya
Fax to 254-20-315977
E-mail scanned copy to: www.knbs.go.ke; director@knbs.go.ke

Access to licensed datasets is only granted when there is a legally registered sponsoring agency (government ministry, university, research centre, national or international organization, etc.).

Requests are reviewed by our data release committee. If approved, you will be provided with the data and documentation on CD-ROM/DVD or through secure FTP server.

Information you provide in this form will not be shared with others, unless a breach to the legal agreement is confirmed in which case the KNBS may inform partner statistical agencies in other countries.

Terms

In this agreement,

1. 'Primary Data Investigator' refers to the investigator who serves as the main point of contact for all communications involving this agreement. The Primary Data Investigator assumes all responsibility for compliance with all terms of this Data Access Agreement by employees of the receiving organization.
2. 'Other Investigators' refers to individuals other than the Principal Investigator, including research assistants, who will have access to the restricted data.
3. 'Receiving Organization' refers to the organization / university / establishment which employs the Primary Data Investigator.
4. 'Representative of the Receiving Organization' refers to an individual who has the authority to represent the Receiving Organization in agreements of this sort.

Section A. Primary Data Investigator

First name

Last name

Title

Prof/ Dr/ Mr/ Mrs/ Ms

Organization

Position in organization

Postal address

Telephone (with country code)

Fax (with country code)

E-mail

Section B. Other Investigators

Provide names, titles, and affiliations of any other members of the research team who will have access to the restricted data.

Name (last / first) Position Affiliation

Section C. Receiving Organization

Organization name

Type of organization

- Line ministry / public administration
- University
- Research centre
- Private company
- International organization
- Non-governmental agency (national)
- Non-governmental agency (international)
- other (specify) _____

Organization website (URL)

Postal address

Section D. Representative of the Receiving Organization

First name

Last name

Title

Prof/ Dr/ Mr/ Mrs/ Ms

Position in organization

Postal Address

Telephone (with country code)

Fax (with country code)

E-mail

Section E. Description of intended use of the data

Please provide a description of your research project (research question, objectives, methods, expected outputs, partners). If information is insufficient, your request may be rejected or additional information will be requested. This information may be provided in an attached appendix to this request.

List of expected output(s) and dissemination policy

Section F. Identification of data files and variables needed

The KNBS provides detailed metadata on its website, including a description of data files and variables for each dataset. Researchers who do not need access to the whole dataset may indicate which subset of variables or cases are of interest. As this reduces the disclosure risk, providing us with such information may increase the probability that the data will be provided.

This request is submitted to access:

- The whole dataset (all files, all cases)
- A subset of variables and/or cases as described below (note that variables such as the sample weighting coefficients and records identifiers will always be included in subsets).

Section G. Data access agreement

The Primary Data Investigator, the Other Investigators, and the Representative of the Receiving Organization agree to comply with the following:

1. Access to the restricted data will be limited to the Primary Data Investigator and Other Investigators.
2. Copies of the restricted data or any data created on the basis of the original data will not be copied or made available to anyone other than those mentioned in this Data Access Agreement, unless formally authorized by the KNBS
3. The data will only be processed for the stated statistical purpose. They will be used for solely for reporting of aggregated information, and not for investigation of specific individuals or organizations. Data will not in any way be used for any administrative, proprietary or law enforcement purposes.
4. The Primary Data Investigator undertakes that no attempt will be made to identify any individual person, family, business, enterprise or organization. If such a unique disclosure is made inadvertently, no use will be made of the identity of any person or establishment discovered and full details will be reported to the KNBS. The identification will not be revealed to any other person not included in the Data Access Agreement.
5. The Primary Data Investigator will implement security measures to prevent unauthorized access to licensed microdata acquired from the KNBS. The microdata must be destroyed upon the completion of this research, unless KNBS obtains satisfactory guarantee that the data can be secured and provides written authorization to the Receiving Organization to retain them. Destruction of the microdata will be confirmed in writing by the Primary Data Investigator to the KNBS.
6. Any books, articles, conference papers, theses, dissertations, reports, or other publications that employ data obtained from the National Data Archive will cite the source of data in accordance with the citation requirement provided with the dataset.

7. An electronic copy of all reports and publications based on the requested data will be sent to the KNBS.
8. The original collector of the data, the KNBS, and the relevant funding agencies bear no responsibility for use of the data or for interpretations or inferences based upon such uses.
9. This agreement will come into force on the date that approval is given for access to the restricted dataset and remain in force until the end date of the project or an earlier date if the project is completed ahead of time.
10. If there are any changes to the project specification, security arrangements, personnel or organization detailed in this application form, it is the responsibility of the Primary Data Investigator to seek the agreement of the KNBS to these changes. Where there is a change to the employer organization of the Primary Data Investigator this will involve a fresh application being made and termination of the original agreement.
11. Breaches of the agreement will be taken seriously and the KNBS will take action against those responsible for the lapse if wilful or accidental. Failure to comply with the directions of the KNBS will be deemed to be a major breach of the agreement and may involve recourse to legal proceedings. The KNBS will maintain and share with partner data archives a register of those individuals and organizations which are responsible for breaching the terms of the Data Access Agreement and will impose sanctions on release of future data to these parties.

Signatories

The following signatories have read and agree with the Data Access Agreement as presented in Section G above:

The Principal Data Investigator

Name _____ Signature _____ Date _____

The Representative of the Receiving Organization

Name _____ Signature _____ Date _____