

**The use of databases of social and gender
statistics in the development of policies and
programmes**

Lancelot Busby



NACIONES UNIDAS



The present document has been prepared by Lancelot Busby, consultant for the Statistics Unit, Subregional Headquarters for the Caribbean, Economic Commission for Latin America and the Caribbean (ECLAC).

The views expressed in this document, which has been reproduced without formal editing, are those of the author and do not necessarily reflect the views of the Organization.

The author wishes to thank Philomen Harrison of the CARICOM Secretariat, Edwin St. Catherine, Director of Statistics in Saint Lucia, Gale Archibald, Project Director in Statistics, Antigua and Barbuda, Corey Gooding of the Statistical Institute of Jamaica and Mc Donald Thomas, Project Officer, Caribbean Development Bank, for their cooperation in granting telephone interviews during the preparation of this document.

Contents

Abstract.....	5
I. Introduction	7
A. Analysis of the current situation.....	7
B. Purpose of the report	7
II. Methodology and conduct of research	9
A. Results of the report	11
1. Type 1 databases.....	11
2. Type 2 databases.....	12
3. Type 3 databases.....	14
III. General appraisal of the three types of database	17
A. The relevance of and extent to which the databases identified are being used	17
1. Type 1 databases.....	18
2. Type 2 databases.....	19
3. Type 3 databases.....	21
IV. Gender databases.....	25
V. An inventory of existing databases of statistics relating to social and gender issues in the Caribbean	27
VI. Conclusions	29
VII. Recommendations	31
Bibliography	33
Annex.....	35
Annex 1 DEVINFO.....	36
Annex 2 NESSTAR.....	37

Annex 3 IHSN	38
Annex 4 SDMX	39
Annex 5 List of persons interviewed.....	40

List of tables

Table 1	Listing of social/gender databases held by Caribbean countries	15
Table 2	Inventory of existing databases of statistics by subject of database.....	28

Abstract

The unavailability of data to inform policy planning and formulation has been repeatedly cited as the main challenge to economic and social progress in the Caribbean. Furthermore, even in instances when data is produced, broader gaps exist between its production and eventual use for evidence-based policy formulation. Owing to those challenges, this report explores the use of databases of social and gender statistics in the development of policies and programmes in the Caribbean subregion. The report offers a general appraisal of databases against two main considerations: (i) maximizing the use of existing databases in relevant policies and programmes; and (ii) bridging the gaps in data availability of relevant statistical databases and their analyses.

The assessment entailed an inventory of social and gender databases maintained by data producers in the region and analysis of the extent to which the databases are used for policy formulation. To that end, a literature search as well as consultations with a number of knowledgeable persons active in the field of statistics and data provision was conducted. Based on the review, a set of recommendations were produced to improve current practices within the region with respect evidence-based policy formulation.

I. Introduction

Caribbean partners have repeatedly noted the insufficient data, inadequacy of analysis and the limited capacity across the Caribbean subregion to provide robust statistical data and information. Governments, among others, need timely and reliable statistics and statistical databases upon which they can base appropriate and solid policy and governance decisions, and build the case for project and programme interventions in any given area. Socioeconomic decisions and gender development investment decisions also rely on the availability of timely databases on social and gender-related issues.

A. Analysis of the current situation

Over the past two decades, international donors and partners have made several contributions towards improvement of regional capacity in the development of social and gender statistical databases. Such assistance has been provided through Caribbean organizations as well as directly to national Governments. Some progress has been made in enhancing the production of social and gender statistics and their accompanying databases. Access to and use of these databases for development purposes, however, remains relatively limited, and even the quantum of available databases is uncertain. It is essential that Member States and Caribbean organizations, to sustain donor interest in providing development assistance, continuously generate and update the relevant statistical databases, and that these be accessible to users as inputs to policies and programmes in the appropriate sectors.

Some committed and dedicated entities have made sterling contributions in this direction. National statistics organizations, such as the Statistical Institute of Jamaica (STATIN) and the Central Statistical Office of Saint Lucia, have been repeatedly cited for their excellence. Additionally, subregional agencies, such as the Caribbean Development Bank (CDB) and the Caribbean Community (CARICOM) Secretariat, have made tangible contributions to the generation of reliable and comprehensive micro databases and in moving the Caribbean subregion towards harmonized approaches.

B. Purpose of the report

The present report is seen to be a positive move contributing to the solution of the lack of good and reliable data that has long characterized Caribbean countries. Two major areas of data paucity stand out.

These are basic statistics, and the organization of available datasets into a format that can be accessed by a wide range of researchers. The current exercise presents a comprehensive report on the availability and use of statistical databases relating to social and gender statistics in the Caribbean subregion and, in particular, on the way these are being used in the development of appropriate policies and programmes in all sectors of development, including the informal sector.

Such a purpose has led to the discovery of the pockets of database effort that exist in the Caribbean. Their lack of coordination is a weakness that, no doubt, has resulted in the continued lack of awareness by many of the database resources that, though limited, are available today and to the asymptotic movement of those databases to excellence. Dialogue, re-design and coordination of effort among the database designers have been observed to be minimal. A low level of analytics, insofar as tracking and analysing demand for data, has resulted in a less-than-satisfactory knowledge of who the users are and what they have done with the data requested. The report abstracts from the observation of all of the databases and databanks and notes the type of collaboration that would be necessary to strengthen the efforts and offerings of each of the database providers.

C. Structure of the report

The contents of this report are structured so that the initial chapter 2 present the findings of the inventory of the main Caribbean databases. Chapter 3 discusses some of the existing databases and databanking initiatives and considers database management systems, resources or platforms that enable the presentation of indicators produced by other databases, and comments on their ability to enable efficient querying of their contents. To the extent that some micro datasets may require delicate handling, practices concerning the release of micro datasets are discussed. In addition, note is taken that some of the hierarchical databases have provisions for the establishment of confidentiality bars.

The section that presents conclusions and recommendations addresses the mathematical and Boolean logic that should underpin all databases on social and gender statistics. The proliferation of database design is observed to render the use of the different databases sub-optimal, as there would inevitably be a long learning curve. The observations made culminate in a set of recommendations for improvement.

II. Methodology and conduct of research

The methodology adopted in conducting the research to fulfil the purpose of the report has taken the form of a four-pronged approach, as follows:

- Telephone interviews and use of email
- Desk review of available and relevant literature
- Conduct of at least one field visit
- Contact with all statisticians at the Advisory Group on Statistics and the Standing Committee of Caribbean Statisticians meeting in Jamaica for purposes of filling in inventory of databases table

Telephone interviews were held with the following (see annex 5):

- a senior representative of the CARICOM Secretariat
- the Director of Statistics in Saint Lucia
- the Project Director in Statistics, Antigua and Barbuda
- a representative of CDB

The interviews were of an unstructured nature, seeking to clarify observations made by the consultant in the course of the research. For example, there was some probing into the existing databases and datasets which should qualify for inclusion in the report. The interviewees were also asked whether they knew of other databases that had not been considered. In the case of Antigua and Barbuda, information was sought on available qualifying databases.

The interviews, both by telephone and through personal visits, discussed the boundaries of datasets that could be considered to be databases of social and gender statistics. There was agreement that the decennial Population and Housing Censuses qualified to be considered as databases. These censuses contained raw, unanalysed facts as collected and, according to the Organisation of Economic Cooperation and Development (OECD) definition, qualified as databanks. By that definition, the Survey of Living Conditions (SLC), the Household Income and Expenditure Survey (HIES), and the Derek Gordon Databank (therein after referred to as “the Databank”), also qualified as databanks.

In addition to the interviews, visits were made to the Databank on the campus of the University of the West Indies in Jamaica, to the PIOJ, and to the office of the Director-General of STATIN in Jamaica. After an initial discussion with the Director-General, the consultant met with an officer of STATIN and was apprised of the European standard for Statistical Data and Metadata Exchange (SDMX2).

The review of databases maintained by PIOJ and STATIN in Jamaica involved a literature search on databases and databanks, and also enabled the consultant to categorize the types of databases that were available in the various countries. The categories, presented in this report, recognize the differences in the databases and datasets collected and made available at country level for access. The meetings with representatives of those two institutions facilitated the preparation of the table detailing the holdings of databases by country. The table did not, however, provide details on the strengths or weaknesses of the data at country office level.

Apart from the interviews, the definitions of database and databank were researched so as to establish what data holdings would be included in this report. The following simple definitions were adopted from the literature:

- Database: A collection of information specific to an operation, business or enterprise.
- Databank: Stored collection of raw, unanalysed facts

More copious discussion on the above terms is available in the OECD Glossary of Statistical Terms 2008. These criteria helped to clarify the two concepts and led to the inclusion in the report of the facilities referred to as databases and databanks.

The report makes an inventory of databases/databanks and describes their contents while observing their mechanisms and, finally, proposing improvements, where applicable.

- Decennial population and housing censuses
- Surveys of living conditions
- Labour force survey statistics
- Multiple Indicator Cluster Surveys (MICS)
- Household income and expenditure surveys
- Education statistics
- And any special surveys on the social situation in any country

The types of social and gender data as housed in the above-named systems/databases are available to different degrees of organization in the Caribbean. The above databases characteristically contain (primary) data at a disaggregated level, collected on the basis of census or sample surveys of households. This means that, depending on the sample design, analysis can reasonably be made at the level of enumeration district or community. The choice of enumeration district or community level will be made on the basis of the location of the issue to be researched (the “development issue”).

One data resource considered its organization of secondary data into a database of indicators to be a useful addition to the stock of data used by researchers, including Government agencies and ministries. Using the criterion of collection of unanalysed data and the requirement of primary data, there is some uncertainty as to whether such a facility is a database in the true sense of the term as defined by the OECD. It may be possible to use the platform to host primary datasets and then, as part of the output, provide value-added products such as indicators – something similar to what Retrieval of Data for Small Areas by Microcomputer (REDATAM) has done in the Caribbean Trade Database (CARIBTRADE), the Merchandise Trade and Transportation Database prepared by ECLAC Subregional Headquarters for the Caribbean.

A. Results of the report

Fundamental to the increased availability of statistics from several databases is their normalization, that gives rise to the ability to join databases for extended access to relevant data, and to the computation of statistics and indicators that could not comfortably be done without the joining of the datasets. Database design that allows for normalization and joining of databases can provide this feature of accommodating joins. It is, therefore, necessary for a measure of design planning, by means of all data suppliers submitting their datasets to organizations such as The Databank or ECLAC for the purpose of establishing regional warehousing and analysis of databases. A controlled amount of redundancy in the storage of national and regional databases is advocated for the purpose of ensured availability of the statistical databases.

It was not immediately clear from the interviews whether the developers of the databases had prepared their datasets in accordance with a common protocol, one that used the same classifications and definitions and the same recording protocols, that would ensure referential integrity and permit joining across datasets. The information gathered permitted the classification of the databases into three groupings. Their description is presented here. The classification was made by the consultant based on the regional outreach and orientation of the databank facility, and the holdings of the databases. For example, population data were considered to be of type 2 if they were national in outlook but with strong collaboration and centralized design.

1. Type 1 databases

a) Observations on preparation of databases by organizations

The nomenclature of the databases into types 1, 2 and 3 is that of the author of the report. Type 1 database is described as being regional in scope, in that it seeks aggressively data from the entire Caribbean subregion. The databases primarily include raw datasets collected from censuses and surveys, and prepared for systematic entry of data into a database-management system. Type 1 is able to systematize any dataset that can be supplied by any official data-producing agency in the Caribbean. This includes national population and housing census data, to the extent that countries make them available to the Databank. The Type 1 database is aggressive in its search for more datasets. It guards the confidentiality of the datasets intensely and ensures that micro datasets are anonymized before being made available to users. This facility takes steps to ensure the bona fides of the persons requesting data and constrains them to state the purpose for which the data will be used. Moreover, it supplies the data on condition that a copy of the report would be provided to the database facility. This is the most organized database observed in the study. The databases in this category are carefully prepared for efficient design and search capabilities. The following data facility falls into the type 1 classification.

Derek Gordon Databank

The mandate of the Databank is to acquire raw numerical data from the Caribbean subregion to facilitate social policy analysis. In addition, the databank provides an archival function because, in the past, many datasets from the Caribbean have been lost through poor archival practices. Although small, the Databank is proud of its collection of datasets, especially from Jamaica. Most of the datasets are from government agencies, but the Databank also welcomes datasets from other organizations/researchers. The utility of Databank is accentuated when one understands that it is aggressive in the collection of raw numerical data from the Caribbean, in most cases without adequate documentation of the datasets being acquired. This places the burden of remedial work on the Databank, to bring the data and documentation to the stage of usefulness to the user.

The Databank receives datasets in a variety of formats. It converts all datasets to Statistical Package for the Social Sciences (SPSS) program format, labels the datasets and carries out consistency checks before distributing the data. Although it stores the datasets in SPSS, it can make the data

available in other software formats. This makes the service of the Databank all the more useful, in that the data can be readily input into a number of packages for further analysis.

The Databank has moved away somewhat from its original adherence to the International Household Survey Network (IHSN),¹ leaning more towards the international Networked Social Science Tools and Resources (NESSTAR) software system, whose main aims are to publish statistics on the website and to combine data and metadata. A Microdata Management Toolkit has been developed by the World Bank Data Group for all online documentation of the datasets towards the NESSTAR system on which the IHSN appears. Included in the toolkit is the NESSTAR Publisher, which is an advanced data management program. It consists of data and metadata conversion- and editing tools, and enables the user to prepare materials for publication to a NESSTAR server. NESSTAR can also be used as a stand-alone tool to prepare data and metadata and save the documentation.

i) Distribution of datasets

The Databank is authorized, with a number of provisos, to distribute unrestricted datasets for academic research. These provisos require that:

- The Databank receive an abstract or a detailed description of any research project using the data
- all users sign an application form
- any payments required are made before datasets are processed
- the user provides the databank with at least one copy of any publication, paper or report produced based on the dataset
- the user informs the databank of any errors in the dataset
- the user does not pass the dataset to other researchers without written permission from the databank
- the user cites the dataset in all reports emanating from use of the dataset

In the case of microdata, the Databank ensures that the data are anonymized, to the extent of not allowing the recognition of any individual through the data provided.

The above stipulations, with the exception of the anonymization, are not often observed.

ii) Depositing a dataset

Depositing a dataset has a number of benefits to the depositors, who can be data-producing organizations such as national statistical offices, central banks, other data producers, or researchers, including university staff. Major benefits to the depositor are the publicizing of their work by the Databank, and the long-term preservation of the data. Guidelines are available for depositors. A data-deposit form should also be completed; otherwise, the Databank will work with potential depositors to obtain the required information.

2. Type 2 databases

a) The decennial population census and household-type survey

Included in the type 2 databases are the national population censuses held by national statistical offices. The census database is perhaps the oldest dataset that has been coordinated across the countries of the English-speaking Caribbean. The censuses fall into the category best described as comprising databases that are national in outlook but with strong collaboration and centralized design. All concepts and

¹ See Mission and Tools of IHSN at Annex 3

definitions have been unified for a core set of questions. The collaboration allows for some flexibility, in that countries can add a module of questions in the census that addresses national concerns that may not necessarily be shared by the other countries. These datasets contain a wealth of data that can be utilized in informing the national planning mechanisms. This is known to governments, yet the data are underutilized. This underutilization is explained by the lack of understanding of the nature of census data. The fact that some data refer to a collection period six years ago does not invalidate their usefulness. Census data, for the most part, are a reflection of behaviours which do not change overnight in response to an economic or natural event, hence their relative stability over time. In some instances of observed change over time, such changes may be attributable to severe socio-economic pressures or natural events including disasters that cause major dislocation in terms of human settlements or livelihoods.

The population datasets facilitate inter-country comparisons in terms of households, individuals, housing, fertility, mortality, morbidity, migration, energy use and a host of indicators of living conditions. The datasets on income are, however, generally not reliable. Several researchers have made use of the unreliable income data to derive Gini coefficients to contribute to their income distribution analyses.

The desire to have greater insights of contemporary social situations leads to the analysis of the population census data by area that approximates the area of occurrence of the situation. This leads to the recognition of communities as an area of interest for continuous analysis. For the evaluation of social policy actions, communities are more relevant than enumeration districts.

For purposes of analysis at any level of the population and geographical hierarchy, Retrieval of Data for Small Areas by Microcomputer (REDATAM), described in greater detail further in the present report, is an excellent example of a database management system for census-type data. REDATAM compares favourably with CPro and SPSS in its ability to analyse population data.

SLC, which is conducted annually by Jamaica but only occasionally in most other countries of the Caribbean subregion, is also included in this category of database. It can be utilized to study a number of gender issues, especially among the poor.

Crime and victimization surveys and migration surveys are conducted by STATIN, and the resulting database is available for query. These two topics are under-researched. The data available at present can go a long way towards the understanding of these phenomena.

b) Multiple Indicator Cluster Survey (MICS) databases

The United Nations Children's Fund (UNICEF) assists countries in collecting and analysing data in order to fill data gaps for monitoring the situation of children and women, through its international household survey initiative, MICS. Access to data is available online through the UNICEF Childinfo webpage. Data collected and offered for access include the following topics:

- Child mortality
- Nutrition
- Child health
- Environment
- Reproductive health
- Child development
- Education
- Child protection
- Survey information

Since the mid-1990s, MICS have enabled many countries to produce statistically-sound and internationally-comparable estimates of a range of indicators in the areas of health, education, child protection, and HIV and AIDS. MICS findings have been used extensively as a basis for policy decisions and programme interventions, and for the purpose of influencing public opinion on the situation of children and women around the world.

c) Basic similarities of type 2 databases

The type 2 databases, as described here, lend themselves easily to the rigorous application of IHSN or NESSTAR for the management of the metadata associated with their data elements. There appears not to have been an initiative aimed at bringing together creators of databases, to agree on a common basic database design that includes the attachment of metadata that provide a useful qualifier to the data being presented.

3. Type 3 databases

a) DEVINFO platform - PIOJ

The type 3 databases are those that are not databases in the sense of their housing raw data and being able to build a table with an engine that incorporates Boolean logic. The type 3 databases are table recall facilities.

JamStats may be considered to be a database if one uses the relaxed definition of a database as discussed earlier in the present report. JamStats captures information on some of the most critical social and economic indicators of Jamaica. Using the DevInfo software developed by UNICEF, STATIN, in partnership with PIOJ, with support from UNICEF, Jamaica, has created a comprehensive database, which allows for the tracking of key development signposts. Integrated with Microsoft Office, JamStats can generate three types of presentations linked to the database - tables, graphs and maps, along with reports that can be customized and formatted as required. JamStats provides data on key socio-economic indicators over multiple time periods and at parish level. Yet, the inability to drill down below parish level limits its usefulness for informing policy, especially at a time when planning must be concerned with a lot more than a macro (nationwide) figure that purportedly speaks about development in Jamaica. The database is therefore seen as non-hierarchical. There is a reality below the parish level that must be understood and factored into social and economic planning. This cannot be offered by JamStats in its present DevInfo design. This limits the usefulness of the tool for the purpose of social and gender planning at community level.

The facility, which is essentially a table recall facility, is useful to a wider community of users – those interested in macro-social and -economic research. The ability to produce a time series is particularly appealing to this type of researcher. Product development must, however, investigate the possibility of incorporating dynamically from other databases, in an inter-operable manner, data elements that can be programmed within DevInfo to calculate the indicator every time that it is requested. In this way, the indicator will stand a chance of reflecting revisions to the numerator, denominator or multiplicand that contribute to the calculation of the indicator. In this manner, the “database” will be able to adjust its output in the context of amendments to its data elements.

In addition to JamStats being produced on the DevInfo platform by PIOJ, the Ministry of Education produces EDUSTATS, which is a “database” on educational statistics. The usefulness of this database is appraised in table 1.

In addition to the telephone interviews, there was a desk review of available and relevant literature. Advice from knowledgeable persons contacted indicated that there was no copious documentation of the existence of functioning databases on social and gender data to inform policy. This suggested that, either there were not many such initiatives, or that these had not been documented. In an e-mail exchange with one colleague, the feeling was transmitted that perhaps students were the major users of the established facility at the University of the West Indies.

A literature search conducted on the Internet yielded important information which has been captured and presented in a number of the annexes to the present report. These annexes describe the holdings of some of the best databases, and proffer them as examples to be kept in mind when Caribbean professionals meet to design and develop databases of excellence.

The table below presents a listing of datasets (loosely termed “databases” as per the discussion earlier in the present report) of social and gender statistics. Many of these datasets (databases) are not accessible to the public at present, thus confining their use to a narrower community of users that would include Government personnel and specially-permitted researchers. The main concern seems to be a certain reluctance to make micro datasets available to a wider range of users, especially if they can be manipulated to identify persons in the communities under research. These datasets (databases) are not housed in the same location and do not benefit from uniform consideration in their storage and protection.

TABLE 1
LISTING OF SOCIAL/GENDER DATABASES HELD BY CARIBBEAN COUNTRIES

Country	Database	Periodicity	Primary or secondary data	Lowest level of geographical representation	Database used by Govt	Database used by others
Antigua and Barbuda	Population Census	Decennial	Primary	ED/Community	Yes	Yes
	HBS	Occasional	Primary	ED/ Community	Yes	Yes
	SLC	Occasional	Primary	ED/ Community	Yes	Yes
	LF statistics	None	n.a.	n.a.	n.a.	n.a.
	ICT	Occasional	Primary	ED/ Community	Yes	Yes
	Education statistics					
	Ad hoc	Annual	Administrative	ED/ Community	Yes	Yes
The Bahamas	Population Census	Decennial	Primary	ED/ Community	Yes	Yes
	HBS	Occasional	Primary	ED/ Community	Yes	Yes
	Survey of Living Conditions	Occasional	Primary	ED/ Community	Yes	Yes
	LF statistics		Primary	ED/ Community	Yes	Yes
	Education Statistics	Annual	Administrative	ED/ Community	Yes	Yes
	Ad hoc					
Barbados	Population Census	Decennial	Primary	ED/ Community	Yes	Yes
	HBS	Occasional	Primary	ED/ Community	Yes	Yes
	SLC	Occasional	Primary	ED/ Community	Yes	Yes
	LF statistics	Quarterly	Primary	ED/ Community	Yes	Yes
	Education Statistics					
		Ad hoc	Annual	Administrative	ED/ Community	Yes
Belize	Population Census	Decennial	Primary	ED/ Community	Yes	Yes
	HBS	Occasional	Primary	ED/ Community	Yes	Yes
	SLC	Occasional	Primary	ED/ Community	Yes	Yes
	LF statistics					
	Education statistics					
		Ad hoc	Annual	Administrative	ED/ Community	Yes
	MICS					
Grenada	Population Census	Decennial	Primary	ED/ Community	Yes	Yes
	HBS	Occasional	Primary	ED/ Community	Yes	Yes
	SLC	Occasional	Primary	ED/ Community	Yes	Yes
	LF statistics	None				
	Education statistics					
		Ad hoc	Annual	Administrative	ED/ Community	Yes
	CWIQ					

(continued)

Table 1 (conclusion)

Guyana	Population Census	Decennial	Primary	ED/ Community	Yes	Yes
	HBS	Occasional	Primary	ED/ Community	Yes	Yes
	SLC	Occasional	Primary	ED/ Community	Yes	Yes
	LF statistics	Quarterly	Primary	ED/ Community	Yes	Yes
	Education statistics					
	Ad hoc	Annual	Administrative	ED/ Community	Yes	Yes
	MICS	Occasional	Primary	ED/ Community	Yes	?
Saint Kitts and Nevis	Population Census	Decennial	Primary	ED/ Community	Yes	Yes
	HBS	Occasional	Primary	ED/ Community	Yes	Yes
	SLC	Occasional	Primary	ED/ Community	Yes	Yes
	LF statistics					
	Education Statistics	Annual	Administrative	ED/ Community	Yes	Yes
Saint Lucia	Population Census	Decennial	Primary	ED/ Community	Yes	Yes
	HBS	Occasional	Primary	ED/ Community	Yes	Yes
	SLC	Occasional	Primary	ED/ Community	Yes	Yes
	LF statistics					
	Education Statistics					
	ICT	Annual	Administrative	ED/ Community	Yes	Yes
	Ad hoc	Occasional				
Suriname	Population Census	Decennial	Primary	ED/ Community	Yes	Yes
	HBS	Occasional	Primary	ED/ Community	Yes	Yes
	SLC	Occasional	Primary	ED/ Community	Yes	Yes
	LF statistics					
	Education Statistics					
	Ad hoc	Annual	Administrative	ED/ Community	Yes	Yes
	MICS					
Saint Vincent and the Grenadines	Population Census	Decennial	Primary	ED/ Community	Yes	Yes
	HBS	Occasional	Primary	ED/ Community	Yes	Yes
	SLC	Occasional	Primary	ED/ Community	Yes	Yes
	LF statistics					
	Education Statistics					
Turks and Caicos Islands	Population Census	Decennial	Primary	ED/ Community	Yes	Yes
	HBS	Occasional	Primary	ED/ Community	Yes	Yes
	SLC	Occasional	Primary	ED/ Community	Yes	Yes
	LF statistics					
	Education Statistics					
	Ad hoc	Annual	Administrative	ED/ Community	Yes	Yes
	Trinidad and Tobago	Population Census	Decennial	Primary	ED/ Community	Yes
HBS		Occasional	Primary	ED/ Community	Yes	Yes
SLC		Occasional	Primary	ED/ Community	Yes	Yes
LF statistics		Quarterly	Primary	ED/ Community	Yes	Yes
Education Statistics						
ICT		Annual	Administrative	ED/ Community	Yes	Yes
Ad hoc		Occasional	Primary			

III. General appraisal of the three types of database

A. The relevance of and extent to which the databases identified are being used

All of the databases listed are relevant, in that they carry data at some level of aggregation/disaggregation that are of interest to a number of researchers, including Governments. Generally, it is the analytics surrounding the use of the data that are not readily forthcoming, for example:

- Who uses the data?
- What datasets are requested?
- How often are requests made by the same person/organization?
- Is there a systematic attempt to request an entire database?
- What are the research papers or articles produced from the data requested?

The Databank tracks the demand for data. This analysis of demand will assist database administrators in improving their offerings and in acquiring those datasets that have been requested but which were not in the possession of the facility.

The extent of use

The disaggregated demand profile for the contents of the database should be continually tracked by relevant record-keeping, as suggested in the discussion of relevance.

The present report comments on the use of the databases according to types 1, 2 and 3 presented in this document.

The report analyses all constraints and their implications relating to the use of statistical databases by database type. The discussion will make use of the headings that examine the following:

- Maximizing the use of the existing databases in relevant policies and programmes

- Bridging the gaps that exist in the subregion with respect to the availability of relevant statistical databases

These headings will be addressed again as the report appraises the different types of database as described.

1. Type 1 databases

Type 1 databases are Caribbean-wide databases that aggressively seek inputs of raw data from censuses and surveys for purposes of organization and storage for search and retrieval.

a) The Databank

The Databank is used quite extensively, as it is located on the Campus of the University of the West Indies at Mona in Jamaica. It has a ready student/researcher clientele. That makes it a relevant and much sought-after resource. There is some indication that it is used by the government planning mechanism, headed by PIOJ, to inform research and decision-making, as well as for programming. The comments concerning its relevance have been made. The facility should, however, prepare analytics to indicate internally the areas that should be tweaked in order to produce an offering of excellence. These are analytics, are listed in the preceding discussion on relevance.

The constraints to the use of the statistical databases at the Databank and their implications

A number of the datasets at the Databank are microdata sets. This imposes a number of constraints on their use. In some cases, the data have to be accessed from a work station within the databank. In all cases of micro datasets that can lead to the identification of any individual, the data must be anonymized before it can be used. This condition is inflexible, and must be enforced so that the suppliers of the datasets can continue to have confidence in the integrity of the databank. In the case of user requests for elements of a database, it is possible that, with systematic querying of a database, the researcher can re-create the database on their computer. The administrators of the Databank stated that a measure of vigilance is exercised by the Databank in this regard, as it keeps track of data requests. The Databank intervenes and stops access if such a motive is detected originating from any systematic query by a researcher. This suggests the possibility of reverse engineering all or part of a database, through data mining. In their text, entitled, “A distance-based approach for database re-engineering”, Barbar and Collard (2001)² focus on the problem of attribute similarity that is quite important for the re-engineering steps according to context-based measures without taking into consideration the naming policy used by database designers

i) Maximizing the use of the existing databases in relevant policies and programmes

The Databank has implemented a strategy to maximize the use of its existing databases, by preparing and disseminating promotional material in the form of brochures. This is seen as a more generalized approach, with all of the disadvantages of dissemination as opposed to data communication. For optimum results, the information concerning the Databank should be targeted to the prospective user-community, to ensure that the message is received. In this way, a higher level of demand is more than likely to follow. The Databank should approach the CARICOM Secretariat for an opportunity to make a presentation of the Databank to the Standing Committee of Caribbean Statisticians, and to

² Barbar, A. Collard, M., published in *Computer Systems and Applications*, ACS/IEEE International Conference , Beirut, Lebanon. 2001

economists, social analysts and project planners, and other professionals. This may help to remove one of the constraints to the fuller use of the facility, namely, insufficient data to input into the databank. Similar presentations, from time to time, to PIOJ and other government officials at the level of permanent secretary may create greater demand for, and use of, the datasets in evidence-based planning. Once success from this approach is secured, the Databank may wish to hold a seminar and invite high-level officials from a number of specific countries to view its operations and discuss submission of their data to the Databank.

The existing databases do not employ the same design and methodology for the storage and retrieval of data. This makes inter-country comparisons of data inexact. Maximized use will require that the data being input into the database be normalized, so that each item number (as in the case of a trade database) has referential integrity (i.e. refers and points to the same data). This requires the assurance that there is a table of equivalence between the data elements of datasets on the same subject matter coming from different countries.

ii) Bridging the gaps in data availability of relevant statistical databases and their analysis

Arising out of the observations at the Databank and the other data facilities studied, it would be desirable for ECLAC and CDB to convene a seminar of the major database constructors to discuss Pan-Caribbean database management systems. This would be a useful follow-up to the “Subregional Seminar on Data Warehousing and Dissemination in the Caribbean” that was organized by ECLAC subregional headquarters for the Caribbean and CDB in November 2010.

The major objective of the seminar was to sensitize Caribbean participants to the need to archive their statistical databases properly for security purposes, in order to facilitate secure data dissemination and use. Hence, there were demonstrations of internationally-accepted methodologies and software, and participants were afforded adequate time to familiarize themselves with the use of those software packages. The software demonstrated has been reported in the present document, after discussions with the main database actors.

Participants reacted positively to the seminar and expressed appreciation for the opportunity to gain greater awareness of the need for collaboration among countries in the Caribbean with regard to data warehousing and dissemination. They recommended the formation of relationships and networks that would facilitate improved communication and dissemination of data among Caribbean countries. The further action recommended by ECLAC and CDB was seen as pertinent.

2. Type 2 databases

a) Population census databases and MICS surveys

The population censuses represent the only dataset or set of databases that are coordinated, in terms of the design and content of the questionnaire and the output tables. In this regard, the selection of software used to process the data is also standardized. Household surveys are closely related to censuses in terms of conduct and processing of the data into analytical tables. To that extent, the current section covers all household survey-type databases. The Survey of Living Conditions and MICS fall into this category, and their attributes and constraints are covered in the subsequent analysis. One major survey of a household nature is the occasional Household Budgetary Survey (HBS) (or Household Income and Expenditure Survey (HIES)). Some comments are included on this type of survey later on in this section relating to type 2 databases.

MICS is a nationally-representative survey of households, women, and children. The main objectives of the survey are to provide up-to-date information for assessing the situation of children and women, and to furnish data needed for monitoring progress toward goals established at the World Summit for Children and to be used as a basis for future action. The fact of its focus on women and children should make the MICS database a prime supplier of gender data to an appropriately-designed database. Its usefulness to the planning mechanism is therefore obvious. The databases – from the

several Caribbean countries that have conducted, and continue to conduct, MICS surveys – are in their infancy, and would benefit from a design that is in keeping with the need for comparability with similar databases from other Caribbean countries. The Databank is poised to perform the transformation of the data into a format for entering the data into the database in Jamaica, using standard data-preparation systems. The official statement on holdings of the Databank conveyed the fact that its major content comprised Jamaican material, although its outreach was Caribbean-wide. In the absence of statistics on use of the databank by researchers outside Jamaica, the notion was that there was not a significant amount of non-Jamaican usage.

The extent of use of the census and household survey databases

These datasets are the most widely used of the social and gender databases in the Caribbean. Extensive use, however, does not imply exhaustive use. There is room for greater use of these databases in facilitating evidence-based decision-making in Caribbean countries. More extensive use of the population and housing census data in databases would result from their storage in one facility, such as The Databank, which remains, arguably, the best organized to handle the assimilation of datasets into its structured system. The Databank can train personnel from other database facilities in the Caribbean on the matter of data handling and control. In terms of demographic data other than the Population and Housing Censuses, administrative records, such as civil registration, population registers and other administrative records, may be pressed into database use. The main problem with administrative records is that, at times, it is difficult to extrapolate totals struck from administrative records to represent the population. At times, assuming complete coverage of administrative data capture, the total from the administrative process should be the national total. This becomes difficult given the incomplete nature of many administrative records.

The Databank is poised to collect and manage data from household surveys. Countries that, at present, do not have the capability for organizing their databases into an efficient databank, should consider the submission of their data to the Databank.

Another useful database management system for the analysis of census and housing data is the REDATAM software of the ECLAC Demographic Centre (CELADE). REDATAM has proven to be effective, as evidenced by its use in countries with large populations. The fact of its being an ECLAC product brings a measure of technical support to Caribbean countries, all of which have census data and would require a degree of security for their datasets. To date, Belize, Saint Lucia, and Trinidad and Tobago make use of REDATAM for this purpose. Saint Vincent and the Grenadines is expected to launch its REDATAM web application soon.

REDATAM as a tool for data processing, analysis and dissemination

CELADE, the Population Division of ECLAC describes REDATAM in these words:³

“It’s a database management tool that administers large volumes of census microdata with hierarchical (geographical) structure, down to the smallest area of the census administrative exercise, often city blocks or similar-sized areas. It was developed to promote access to, and analysis of, census and other data, to inform decision-making for sectoral and local development programmes and policies – by facilitating the dissemination by NSOs of microdata (taking into account confidentiality issues) – by providing the end users (Sectoral Ministries, regional and local authorities and universities) with a user-friendly, fast software.”

³ Expert Group Meeting on Affordable Technologies for Disseminating Official Statistics, Bangkok, 5-7 October 2005

b) The Household Budgetary Survey (HBS) or The Household Income and Expenditure Survey(HIES)

The Household Budgetary Survey, otherwise called The Household Income and Expenditure Survey, is occasional and serves to update the knowledge base on incomes and consumer behaviour. Its main direct use is the derivation of items and weights for the revision of the Consumer Price Index, or the Retail Price Index, as some countries term it.

Caribbean Household Surveys Databank (CHSD)

CHSD was created by ECLAC under the project Improving Household Surveys in the Caribbean, which was financed by the State Department of the United States. The Databank was built to facilitate electronic access to and dissemination of systemized micro-level data via the Internet. It features data sets on households from a range of Caribbean countries and on a range of issues including household income and expenditure, from 1990 onwards. For each dataset, the databank allows access to user manuals and metadata on the survey, along with the micro-level dataset. Users have the option of generating tables or frequencies based on their specifications. Restrictions imposed by Governments and statistical agencies on the use of the household survey datasets means that access to the databank is restricted and can only be authorized upon request. This would serve as a natural limitation on its widespread use.

A high-level presentation to governments through the ministries of planning and social development, together with the respective national statistical offices and special interest organizations, such as UN WOMEN and the United Nations Population Fund (UNFPA), that are interested in issues such as gender equality, domestic violence and women's empowerment, should result in the production of the required datasets that would bridge the statistical database gap that exists in the Caribbean. Such an initiative may well result in the greater supply of indicators to address the goals and targets of the Millennium Development Goals.

Analysis of constraints and their implications, relating to the use of statistical databases e.g. accessibility

Most Caribbean countries do not yet conduct inter-censal surveys on a continuous basis hence population census databases outnumber those resulting from household surveys. Within more recent times, CDB has spearheaded a multi-donor initiative called Support to Poverty Assessment and Reduction in the Caribbean and this initiative has assisted to enhance the capability of the Organisation of Eastern Caribbean States. A central repository for these datasets will help to remove constraints to their use but the constraint of confidentiality remains.

i) Maximizing the use of the existing databases in relevant policies and programmes

Maximizing the use of the existing databases is closely co-related to both their availability and their design. Appropriate design would facilitate their use. Bridging the gaps that exist in the subregion with respect to the availability of relevant statistical databases

A number of good initiatives to produce databases and survey capabilities have been made by agencies such as CDB, the United Nations Development Programme (UNDP), UNICEF, and ECLAC through its efforts to build a CHSD based on the Banco de Datos de Encuestas de Hogares (BADEHOG) model of ECLAC, Santiago..

3. Type 3 databases

The databases classified as type 3 in the present report are essentially those that are designed on the DevInfo platform. This platform, while useful for quick access to indicators, is fundamentally

weakened by its capture of indicators as opposed to the component statistics that may be used to calculate the indicators. The datasets affected by the DevInfo design are mainly:

- Economic statistics
- Social statistics, that include health and education statistics.

This platform is not as rigorous as the relational or hierarchical database management systems. It may, however, be possible for some work to be done to provide an element of hierarchical structure to the platform, so that it can produce analyses of social situations at different levels of geographical or social aggregation or disaggregation. If such a re-design can be done, the type 3 databases will be even more effective than they are at present.

a) Detailed account of all type 3 databases and the extent to which they are being used

The DevInfo Databases are present in all Caribbean countries, in fulfilment of the original design. They are used extensively by ministries of planning and social development, and to track the Millennium Development Goals. Any improvement to their design that would make for a sturdier data recall facility would be welcome.

b) Analysis of all constraints and their implications, relating to the use of statistical databases e.g. accessibility

The design constraints of this type of database (type 3) have been commented on earlier in the text of the present report. A hierarchical database would yield more information and, with the use of some of the present capabilities of DevInfo, produce value-added analytical data at more levels than is now possible. The accessibility of this type of database would be enhanced if it were to feed into a regional repository. In the present form, the maintenance of these databases would result in considerable traffic between the constructors of the databases and the repositories. One alternative would be that, in the interest of referential integrity, a request coming in to repositories such as the Databank or ECLAC would be redirected to the national centres where this type of database is held, and receive from those centres a response that can be passed back to the researcher making the request.

Maximizing the use of the existing databases in relevant policies and programmes

The databases under discussion in the current section of the report are, without doubt, useful. Addressing the constraints discussed above will make the data more useful. As in the case of the Databank, a policy of stating, at the time of requesting the data, a commitment to providing the Databank with a copy of the resulting paper, could well be a catalyst to the more widespread use of the facility. If the repository, in cataloguing its holdings, were to publish abstracts of the documents produced using the holdings of the facility, this would, in all probability, lead to greater use of the facility in response to more aggressive promotion of the facility through the published abstracts. Aggressive marketing of the holdings and uses of the databases is a desirable new stance that should be taken by providers of data. It is possible that there exists a segment of organizations and people unaware of the benefits to be derived from use of these databases. If this were the case, the owners of the databases could showcase their resources and demonstrate in sales seminars the ways in which the data could be used to add value to a range of organizations, using actual situations and examples of the use of the data. Maximization of the use of existing databases should not be viewed as being predominantly the domain of the public sector. Market intelligence is fuelled by data emanating from population and housing censuses and other household surveys.

Bridging the gaps that exist in the Caribbean subregion with respect to the availability of relevant statistical databases

The observed reactive stance of the repositories for data must be replaced by service orientation. To this extent, the Databank and other producers of databases should showcase their products and services from time to time, maybe in a designated “Data Accessibility Week”. The need for a database design seminar, as discussed above, is seen as a way through which an extension to the data that can be accessed from datasets operating in isolation can be realized from the joining of databases. This would require a degree of redesign of systems. The recommended ECLAC and CDB initiative may achieve gains in this respect. A number of organizations with interest in specific aspects of social statistics may wish to be a part of a funding arrangement that would see:

- the emergence of new databases such as those on gender, that would include databases on domestic violence, women in government, and the contribution of female hucksters and the informal sector in general, to the national economy
- demonstrations of the use of downstream analysis using specialist analytical tools such as STATA and SPSS (now PASW)
- the link between analytical tables and thematic maps as a most persuasive communicator of the development problem to be addressed.

The gaps that exist at present are as much attributable to systems as they are to people. Concerted action is seen as a powerful way to make gains in the provision of coherent, relevant and quality social and gender statistics in the Caribbean.

IV. Gender databases

Investigations on the databases in existence yield the notion that there are no databases on gender. The sex data afforded by the household surveys are not, correctly speaking, databases on gender. Gender is not the same as sex. A gender database is expected to carry data on the issues that occupy the attention of UN Women and other associated special-interest organizations. A gender database should therefore incorporate datasets on:

- Violence against women
- Peace and security
- Leadership and participation
- National planning and budgeting
- Economic empowerment

In this context, the input of UN Women, UNFPA, UNICEF and other like-minded organizations would be valuable to the design and construction of an excellent, Caribbean-wide gender database, with its country chapters that contribute in a uniform manner to the Caribbean facility.

ECLAC, in collaboration with UN Women and UNDP, conducted a gendered analysis of the impact of Hurricane Ivan on Grenada in 2005 using the Core Welfare Indicators Questionnaire approach. This analysis addressed a number of gender issues, but was an ad-hoc exercise and was not considered to be a database for the purpose of the present report. The design of a gender database can, however, benefit from the ECLAC study on Grenada.

A more recent development has been the convening of a subregional meeting on “Enhancing the Capacity of Caribbean Countries to Eradicate Violence against Women.” This forum was held in Saint Lucia on 15 June 2010. The meeting discussed the role of statistical data and indicators in the monitoring of violence against women as a prerequisite to establishing policies to eradicate such violence. The meeting considered the role that administrative registries could play in the measurement of violence against women and looked at the present efforts to collect data on domestic violence. The meeting concluded that:

- standardization of the approach to the measurement of domestic violence should be accomplished

- data gathered needed to cover both perpetrator and victim
- use could be made of administrative records, but that this did not remove the need for subject-specific surveys to be conducted from time to time
- collaboration among all Caribbean countries was needed and all stakeholders needed to be trained in the collection of data on domestic violence.

The major observation emanating from the meeting confirmed the earlier statement on the infancy of data gathering and analysis of statistics on domestic violence, in particular, and gender statistics in general.

In view of the foregoing, it would be safe to say that there are no databases on gender statistics on the Caribbean at the present time.

V. An inventory of existing databases of statistics relating to social and gender issues in the Caribbean

The contacts to guide access to, and use of, the databases are provided by reference to the column that refers to location of data. The ratings in the last column, that address apparent quality of data, were not based on direct answers. The ratings are, to a great extent, derived from the familiarity of the consultant with the data and from perceptions gathered from discussions with knowledgeable persons. For instance, the rating given to education data from the ministry of education is informed by some familiarity with the data-collection processes in several countries.

In keeping with the definition of database presented earlier, there is a housing database, but its quality may not be excellent. This database, or some aspect of it, may reside with the national statistical office in its census database. In addition, the ministry responsible for town and country planning possesses some data on housing, mainly derived from its administrative processes. These datasets considered cannot be described as anything but fair, at most.

To the extent that the census records in the past have recorded responses to the questions of income, and to the extent that these questions have been processed, one may argue that there is a database on remuneration. There is a widely held view that the quality of the responses to the income questions is poor. The answers to the income question are thought to be conditioned by the feeling on the part of the respondents that their answers will be used for income tax assessments. That does not negate the existence of the database. The poor quality of the data may result in the responses to the income question not being used. The table of this database has been included to emphasize its desirability and to keep it in the accounting.

TABLE 2
INVENTORY OF EXISTING DATABASES OF STATISTICS BY SUBJECT OF DATABASE

SUBJECT OF DATABASE	COUNTRIES	SOURCE	LAST DATE / PERIODICITY	LOCATION OF DATA	APPARENT QUALITY OF DATA
Population	All countries	Population Census	2000 / 2001	NSO, DCDB (some)	Satisfactory to good
Education	All countries	Ministry of Education admin. records	2008?/ Annual	Ministry of Education	Satisfactory to good
Health	All countries	Ministry of Health	2008?/ Annual	Ministry of Health	Fair
Housing	All countries	Ministry of Housing / Planning	?	NSO, Ministry of Housing / Planning	Fair
Employment	Trinidad/Tobago Jamaica Barbados Guyana	Household Labour Force Survey	2008?/ Annual	NSO Ministry of Labour	Fair
Remuneration	All countries	Population Census Survey of Living Conditions	2008?/ Annual	Population Census SLC	Poor
Poverty	All countries	Population Census Survey of Living Conditions	Varying dates / Occasional	NSO, Ministry of Planning Ministry of Social Affairs	Fair
Income Distribution	All countries	Population Census Survey of Living Conditions	2000 / 2001 Varying dates / Occasional	NSO	Fair to poor
Gender	All countries	Limited Census data	2000 / 2001	NSO	Poor. Not particularly gender-oriented
Miscellaneous Indicators	All countries	National Statistical Office, Planning Institute of Jamaica	2008?	NSO, PIOJ	Fair to good
Some of all of the above data	All countries	The Databank	Varying datasets and periodicity	Derek Gordon Databank	Good with good effort

VI. Conclusions

The conclusions presented in this report derive from the terms of reference and the outputs desired from the research.

One conclusion is that there is fragmentation in the conceptualization, construction and management of databases in the Caribbean. The social and gender statistics databases reflect this lack of coordination and networking between database constructors. This is as true for both the national and regional database landscapes.

The design of a database is grounded in mathematics that takes the designer into the realms of set theory, categorization, normalization and Boolean logic, among other areas. It is necessary that, before more ground is covered by the designers of statistical and other databases, there should be a forum at which the fundamentals of databases and information systems are fully discussed and adopted into the database constructors' mindset. This will allow some convergence in the design of statistical databases. The forum should be able to enunciate a narrower definition of the term "database" that would set the tone for conceptualization and design of information systems.

In order to secure referential integrity across datasets, the concepts, definitions and codification schemes should be unified, to the extent possible, in order to avoid the need for a proliferation of re-codes, many of which will not yield one-to-one correlations with other classification schemes. This means that analytical comparative tables produced will be unnecessarily impaired by proximate datasets. Data handling is an area of expertise that does not seem to be prominent and fundamental in the case of national databases. The Databank has demonstrated a keen understanding of the need to be meticulous in data handling, and should lead the discussion on this matter at the recommended forum of database designers. Similarly, REDATAM requires meticulous handling of data to be input into the database. The decision of Caribbean countries to adopt REDATAM will lead to the creation of a common mindset towards the construction and use of databases.

Many of the social statistics databases are hierarchical in nature, are derived from household surveys that employ similar sampling and data processing paradigms, and coding schemes that are similar or can be easily unified throughout the region. There is a prime software platform designed by ECLAC in association with the University of Waterloo and vastly improved by the expert intervention of the late Serge Poulard. This is the REDATAM software that is hierarchical in nature. Its original intent was to enable the processing and retrieval of data for small areas. It is ideal for the creation of databases on census and survey data. A number of years ago, ECLAC Subregional Headquarters for

the Caribbean designed and constructed a trade statistics database using the REDATAM platform, with great success. It is a value-added database, that in addition to the retrieval of data and the building of tables in response to queries (including online queries from remote locations), computed indicators in a number of areas of concern upon request. The search engine is extremely fast, in contrast to its earlier versions, and it is capable of processing census data from very large populations (Brazil being just one example of a large population). In the construction of the database, metadata was an area of concern and REDATAM provided a useful solution. Several years later, the World Bank IHSN and NESSTAR metadata toolkit initiatives emerged, as well as the series of Statistical Data and Metadata eXchange (SDMX) initiatives. It is possible that collaborative work inspired by an ECLAC initiative could result in the incorporation of one of these toolkits into the design of REDATAM, to yield a near-perfect database format to be adapted marginally in accordance with the nature of the data being stored, and to be, perhaps, the stock database design on which many now non-existent databases can be built at national level across the Caribbean. One can think of the incisive tables at varying levels of geography, income status and community that can be derived to throw light on the issues identified from the perspectives identified here.

REDATAM, as designed for the CARIBTRADE database, also incorporates access bars that effectively delimit the extent to which drilling down would be permitted to certain identified groups. This is effective protection against over-disclosure that can lead to the identification of the particulars of certain establishments that could lead to the identification of other aspects of their profitability and bottom line. Other techniques of anonymization may be added to REDATAM, or to its output tables, that can take the form of a number of file formats.

CARIBTRADE has not been updated because of data supply problems, as countries are reluctant to give to ECLAC data that they give to the CARICOM Secretariat. Their preference is that CARICOM and ECLAC would share the data.

The first observation is that there has not been an integrated approach to the development of databases in the Caribbean. This has resulted in the use of a number of database platforms that tend to make the learning curve long because of the numerous packages that have to be learnt.

Another observation is in respect of the type 3 databases. These are useful in the publication of indicators. If DevInfo can be regarded as being a “front end” to one of the more narrowly-termed “databases,” it could remove its major weakness of duplication of data entry by calculating indicators at different levels of the data hierarchy, and publishing them. If DevInfo can be redirected to this type of function, it would constitute a value added product, way in excess of its current, acknowledged value added attributes.

One conclusion is, therefore, that no one database reviewed seems to be the perfect database. The best features of the separate databases should be incorporated into a new design that should guide database development work in the Caribbean.

A second conclusion is that gender databases are almost non-existent. There is an expressed desire to have them constructed. However, the specialist gender-oriented organizations may not have the time or other resources to construct them. They can, however, participate in the design of these datasets.

The third conclusion presented here is that it would be better for the constructors of databases to utilize the same concepts, definitions and coding schemes, as far as possible, to facilitate national, regional and international comparisons.

VII. Recommendations

The conclusions discussed above lend themselves easily to recommendations. The recommendations are the following:

- ECLAC, in collaboration with CDB, should convene a series of technical meetings to discuss the data gaps in the area of social and gender databases in the Caribbean, and plan an integrated approach to their construction, using, as far as possible, the same database platform for storage of the raw data
- The partners should examine the available toolkits for the treatment of metadata and adopt one in the interest of greater universal use and understanding. The Databank should be a leader in the discussion of the treatment of metadata.
- The partners should examine the social and gender statistical databases of the United Nations and ECLAC Santiago, to gain a rapid appreciation of the content and design of those databases, so as to move ahead with the benefit of other organizations' experiences.
- The partners should acknowledge the importance of using the same concepts, definitions, classification and coding schemes in the interest of referential integrity as well as notional and regional comparability.
- PIOJ should lead the discussion of the possibility of redesigning DevInfo to more comfortably accommodate hierarchical datasets, and produce geographically and otherwise-layered, tabular results.
- STATIN should lead the discussion on gender-based statistical databanks, and showcase its study on Violence Against Women. At that session, UN Women should make a presentation on the data it would like to capture in such a database.

Bibliography

- Barbar, A. and Collard, M. (2001), in *Computer Systems and Applications*, ACS/IEEE International Conference, Beirut, Lebanon
- UNESCAP, (2005), Expert Group Meeting on Affordable Technologies for Disseminating Official Statistics, Bangkok, 5-7 October 2005
- ECLAC, (2010), Evaluation Report of the Subregional Seminar on Data Warehousing and Dissemination in the Caribbean, ECLAC LC/CAR/L.285
- ECLAC, (2011), Report of the Subregional Meeting on Enhancing the Capacity of Caribbean Countries to eradicate violence against women, ECLAC LC/CAR/L.297
- Thomas Davenport, Cohen and Jacobson, (2005), *Competing on Analytics*, Babson Executive Education, Babson Park, MA USA

Annex

Annex 1 DEVINFO

facts as presented in website URL

<http://www.devinfo.org/>

What follows is a direct quote from the URL

DEVINFO

DevInfo is an integrated desktop and web-enabled tool that supports both standard and user-defined indicators. The standard set of MDG indicators is at the core of the DevInfo package. In addition, at the regional and country levels, database administrators have the option to add local indicators to their databases. The software supports an unlimited number of levels of geographical coverage: from global level to regional, sub-regional, national and sub-national down to sub-district and village levels (including schools, health centers, water points).

DevInfo has simple and user-friendly features that can be used to query the database and generate tables, graphs, and maps. The system provides an ideal tool for evidence-based planning, results-focused monitoring, and advocacy purposes. It allows for organizing, storing and displaying data in a uniform way to facilitate data sharing at the country level across government departments, UN agencies, and development partners.

Data from DevInfo can be exported to XLS, HTML, PDF, CSV and XML files and imported from spreadsheets in a standardized format. DevInfo also has a data exchange module for importing data from industry-standard statistics software packages such as SPSS, SAS, Stata, Redatam, and CSPro.

Annex 2 NESSTAR

Advanced Data Management

This description is taken directly from the following url:

<http://www.nesstar.com/software/publisher.html>

The following paragraphs have been selected from the text on the DevInfo Website referred to above to describe its main characteristics.

Nesstar Publisher is an advanced data management program. It consists of data and metadata conversion and editing tools, enabling the user to prepare these materials for publication to a Nesstar Server. However, it can also be used as a stand-alone tool for the preparation of data and metadata. The Publisher enables users to enhance datasets by combining a wide range of catalogue and contextual information, which can then be viewed within the Nesstar web client, Nesstar WebView.

A complete metadata authoring tool

The key feature of Nesstar Publisher is that it is a complete metadata authoring tool. Metadata is critical because data are only made accessible through their accompanying documentation. Without a description of their various elements, data resources will manifest themselves as more or less meaningless collections of numbers to the end user. The metadata provide the bridge between the producers of data and their users and convey information that is essential for secondary analysis.

DDI compliant

Nesstar Publisher is also DDI compliant. The DDI (Data Documentation Initiative) is a metadata standard used for documenting datasets developed in European and North American data archives, libraries and official statistics agencies. It is designed to be fully machine readable and machine processable. The DDI is now defined in XML which facilitates easy Internet access but the advantage of Nesstar Publisher is that the user does not need to have any knowledge of XML to generate data documentation.

Annex 3

IHSN

The following text has been adapted from the International Household Survey Network website that can be accessed at the following URL: <http://www.surveynetwork.org/home/>

Microdata Management Toolkit

The Microdata Management Toolkit developed by the World Bank Data Group for the International Household Survey Network aims to promote the adoption of international standards and best practices for microdata documentation, dissemination and preservation.

The Toolkit comprises three modules. The **Metadata Editor** is used to document data in accordance with international metadata standards (DDI and Dublin Core). The **Explorer** is a free reader for files generated by the Metadata Editor. It allows users to view the metadata and to export the data into various common formats (Stata, SPSS, etc). The Metadata Editor and Explorer are based on Nesstar technology and developed by the Norwegian Social Science Data Services (NSD). The **CD-ROM Builder** is used to generate user-friendly outputs (CD-ROM, website) for dissemination and archiving.

Objectives

The mission of IHSN is to foster the improvement of the availability, accessibility and quality of survey data in developing countries, and to encourage their analysis and use by national and international development decision-makers, the research community and other stakeholders.

Annex 4

SDMX

This text is taken from a description of SDMX that is presented on the website of the Statistical Data and Metadata Exchange. It can be referenced at http://sdmx.org/?page_id=7

What is SDMX?

The Statistical Data and Metadata eXchange refers to an international initiative aimed at developing and employing more efficient processes for the exchange and sharing of statistical data and metadata among international organizations and their member countries. The initiative, started in 2001, is sponsored by seven international organizations: Bank for International Settlements (BIS), European Central Bank (ECB), Eurostat, International Monetary Fund (IMF), Organisation for Economic Co-operation and Development (OECD), the United Nations and the World Bank (WB), who are committed to establishing, implementing and complying with common standards.

The SDMX version 1.0 set of technical standards has been approved by ISO as a technical specification (ISO/TS 17369:2005). Version 2.0 has been publicly released in November 2005 and its approval process by ISO is ongoing.

What are the goals of SDMX?

The rationale of SDMX is standardization for statistical data and metadata access and exchange.

With the ever-increasing ease of use of the Internet, the electronic exchange and sharing of data is becoming more and more easy, frequent and important. This stresses the need for a set of common standards for exchange and sharing of statistical data and metadata, and for making processes more efficient. As statistical data exchange takes place continuously, the gains to be realized from adopting common standards are considerable, both for data providers and users.

The objective is to establish a set of commonly-recognized standards to which all players adhere, making it possible to have easy access not only to statistical data, wherever these data may be, but also to metadata that make the data more meaningful and usable. The standards will allow national organizations to fulfil, in a very efficient way, their responsibilities towards users and partners, including international organizations by, among other things, using their general online databases to give access to data as soon as they are released.

The SDMX standards also aim to ensure that appropriate metadata always come along with the data, making the information immediately understandable and useful. For this reason, standards for metadata exchange are extremely important in SDMX.

When can SDMX be used?

As mentioned above, the SDMX standards are designed for exchange or sharing of statistical information between two or more partners. Evidently, the SDMX standards have been developed by the sponsors in order to accommodate the constituencies of the sponsoring organizations (national statistical offices, central banks, ministries, etc.). Within and across these constituencies, the standards are intended for reporting (or sharing) statistical data and metadata in the most efficient way.

SDMX standards can also be used within a national system for transmitting or sharing statistical data and metadata, and by private data providers (such as re-sellers of statistical databases). This is particularly interesting in countries with a federal structure or a fairly decentralized statistical system. In such cases, a close link can be established between the national system for data sharing and the international ones, allowing for additional efficiency gains for the organizations involved.

If data are made available for exchange using the pull mode (see Box 1) according to SDMX standards, this could easily evolve to open SDMX-based dissemination; such dissemination may respond well to user demands for well-structured data and metadata in reuseable formats, and should be considered as an option for national authorities as well as international organizations.

Annex 5

List of persons interviewed

- Philomen Harrison, Caribbean Community Secretariat
- Edwin St. Catherine, Director of Statistics, Saint Lucia
- Gale Archibald, Project Director in Statistics, Antigua and Barbuda
- Mc Donald Thomas, Project Officer, Caribbean Development Bank
- Kristin Fox, Manager, Derek Gordon Databank
- Richard Leach, Database Administrator, Derek Gordon Databank
- Sonia Jackson, Director–General, the Statistical Institute of Jamaica
- Caren Nelson, Manager, Policy Research Unit, The Planning Institute of Jamaica
- Donneth Edmondson, Director, JAMSTATS Secretariat, The Planning Institute of Jamaica
- Collette Robinson, Manager, Social, Welfare and Gender, The Planning Institute of Jamaica
- Frederick Gordon, Programmer, SLC, The Planning Institute of Jamaica
- Corey Gooding, the Statistical Institute of Jamaica