



Economic Commission for Latin America and the Caribbean  
Subregional Headquarters for the Caribbean

---

Expert group meeting to examine the potential for expanding  
access to Caribbean census microdata for researchers  
23 November 2015  
Port of Spain, Trinidad and Tobago

LIMITED  
LC/CAR/L.484  
17 December 2015  
ORIGINAL: ENGLISH

## **REPORT OF THE EXPERT GROUP MEETING TO EXAMINE THE POTENTIAL FOR EXPANDING ACCESS TO CARIBBEAN CENSUS MICRODATA FOR RESEARCHERS**

---

This report was reproduced without formal editing.



## CONTENTS

A.	CONCLUSIONS AND RECOMMENDATIONS .....	1
B.	ATTENDANCE AND ORGANIZATION OF WORK.....	2
1.	Place and date of the meeting.....	2
2.	Attendance .....	2
3.	Documentation .....	2
4.	Agenda .....	2
C.	SUMMARY OF PROCEEDINGS .....	2
1.	Opening of meeting .....	2
2.	Objectives of the meeting.....	3
3.	Presentation of the report “Dissemination of Caribbean census microdata to researchers including an experiment in the anonymization of microdata from Grenada and Trinidad and Tobago” .....	3
4.	Data archives: Partners in expanding access to microdata.....	5
5.	Closing remarks.....	7
Annex I	List of participants .....	8

## A. CONCLUSIONS AND RECOMMENDATIONS

1. All participants agreed that steps should be taken to extend access to Caribbean census microdata.
2. Of the various approaches considered, remote access was seen as potentially the most fruitful for small island States since researchers could be provided with controlled access to full count microdata rather than samples of records as is often the case with public and licensed use files. The Minnesota Population Center (MPC) at the University of Minnesota, through the IPUMS-International project, plans to pilot remote access facilities in the near future using census data for Trinidad and Tobago.
3. Samples of anonymized census records disseminated as either licensed or public use files can be analytically useful for the larger Caribbean countries (Jamaica, Trinidad and Tobago, Guyana and Suriname); they could be of some marginal utility for countries with populations of around 300,000 (Bahamas, Barbados and Belize); and are of limited value for countries with smaller populations. Release of licensed use files, rather than public use files, is likely to be an easier first step for countries with no previous experience of microdata release since licensing provides an extra layer of protection against the risk of disclosure.
4. There were mixed views with regard to the viability of secure data laboratories that require researchers to visit the laboratory on-site to access microdata. Some participants questioned whether microdata laboratories would attract sufficient users to make them sustainable. The statistical offices of Belize and Suriname currently offer on-site access to microdata and an analysis of user traffic at these facilities would offer insights to the viability of secure data laboratories in the Caribbean.
5. Caribbean statistical acts do not explicitly authorize microdata release although they do not prohibit it. Modernising statistical legislation to cover arrangements for microdata release would facilitate greater access to microdata by formally recognising this important function of statistical offices and providing a legal basis for data access agreements.
6. The finalized study, "Dissemination of Caribbean census microdata to researchers including an experiment in the anonymization of microdata from Grenada and Trinidad and Tobago", should be disseminated more widely to National Statistical Offices (NSOs) in the Caribbean subregion. It was also suggested that the document could be of interest to those responsible for dissemination of census data in other regions, especially in the Pacific.
7. ECLAC is already actively involved in promoting the widest possible use and dissemination of census data through its support for, and promotion of, the REDATAM software. ECLAC will continue to support and promote the REDATAM software but will also seek to encourage and support statistical offices in other forms of microdata dissemination.
8. Member States should consider depositing census data with both IPUMS-International and the Derek Gordon Databank at The University of the West Indies (UWI), Mona Campus.
9. IPUMS-International and the Derek Gordon Databank will collaborate with each other and NSOs to archive census microdata, to expand access to Caribbean census microdata for researchers, and to protect the confidentiality of census records.

## **B. ATTENDANCE AND ORGANIZATION OF WORK**

### **1. Place and date of the meeting**

10. The expert group meeting to examine the potential for expanding access to Caribbean census microdata for researchers was held on 23 November 2015 at the Economic Commission for Latin America and the Caribbean (ECLAC) subregional headquarters for the Caribbean, Port of Spain, Trinidad and Tobago.

### **2. Attendance**

11. The meeting was attended by representatives of the Central Statistical Office of Grenada, the Statistical Institute of Jamaica, the Statistical Office of Saint Vincent and the Grenadines and the Central Statistical Office of Trinidad and Tobago. A representative of The United Nations Population Fund (UNFPA) was present. Also present were representatives of the Minnesota Population Center at the University of Minnesota (coordinators of the IPUMS-International project), and the Derek Gordon Databank, part of the Sir Arthur Lewis Institute of Social and Economic Studies, located at the Mona campus of the University of the West Indies, Jamaica. Representatives of the Caribbean Community (CARICOM), the World Bank and the Central Statistics Office of Saint Lucia were unable to participate in the meeting but submitted written comments on the study which are reflected in this report.

### **3. Documentation**

12. The study “Dissemination of Caribbean Census Microdata to Researchers including an experiment in the anonymization of microdata from Grenada and Trinidad and Tobago”, prepared by ECLAC subregional headquarters for the Caribbean served as the main background document for the meeting.

### **4. Agenda**

1. Opening of the meeting
2. Objectives of the meeting
3. Presentation and discussion of the paper “Dissemination of Caribbean Census Microdata to Researchers including an experiment in the anonymization of microdata from Grenada and Trinidad and Tobago”
4. Presentations and discussion: ‘Data archives: partners in expanding access to microdata’
5. Conclusions and recommendations
6. Closing remarks

## **C. SUMMARY OF PROCEEDINGS**

### **1. Opening of meeting**

13. Welcome remarks were made by the Coordinator of the Statistics and Social Development Unit on behalf of the Director of ECLAC subregional headquarters for the Caribbean. He conveyed the

importance of exploring ways to expand access to microdata in the Caribbean subregion while protecting statistical confidentiality.

14. He noted that often, the real problem in the Caribbean is inadequate dissemination of data, rather than simply paucity of data. The purpose of the meeting was to explore the issues surrounding access to census microdata in the Caribbean. He added that the ultimate goal was to provide access to microdata to promote the use of census data for social research, development planning and policymaking.

15. The Coordinator of the Statistics and Social Development Unit acknowledged that the small population sizes of Caribbean Member States made it more difficult to anonymize census microdata. He expressed the hope that the ECLAC study, which would be presented to the meeting, would contribute to overcoming this challenge. He urged the meeting to discuss the findings of the paper and identify concrete steps to expand access to census microdata.

## **2. Objectives of the meeting**

16. The Population Affairs Officer of the Statistics and Social Development Unit of ECLAC subregional headquarters for the Caribbean stated that there were two main objectives for the meeting. The first objective was to get the comments and feedback of meeting participants on the ECLAC study. The second objective was to engage in an exploratory discussion on practical steps that could be taken by ECLAC, other international organizations and data archives such as IPUMS-International and the Derek Gordon Databank in order to open up access and expand availability of Caribbean census microdata to researchers in the Caribbean subregion and worldwide.

## **3. Presentation of the report “Dissemination of Caribbean census microdata to researchers including an experiment in the anonymization of microdata from Grenada and Trinidad and Tobago”**

17. The Population Affairs Officer of ECLAC presented the study to the meeting. He outlined the challenges facing Caribbean statistical offices in the dissemination of census microdata, described some commonly used disclosure control methods for census data, and discussed their applicability to the Caribbean. He identified the main challenges to include the small size of Caribbean countries which makes data anonymization more difficult, limited human resource capacity in statistical offices, outdated statistical legislation which does not provide a governing framework for microdata release, and the relatively small number of Caribbean researchers interested in accessing census microdata.

18. He described three levels of data dissemination: traditional published census reports, interactive tabulation tools such as REDATAM, and the release of microdata. The presentation outlined the technical, administrative, and legal procedures that are commonly used to release census microdata. He described how these different types of procedures are used in combination to release more or less heavily protected microdata to different categories of user. He noted that for each category, a balance has to be struck between protecting statistical confidentiality while maximising, as far as possible, the analytic utility of the microdata that is released.

19. He then reported on an experiment to apply disclosure control methods to the census data of Grenada and Trinidad and Tobago. The Population Affairs Officer described how disclosure risk was analysed, and how this informs the design of microdata release files, whether they are released as public use files or licensed use files. He also highlighted the limitations of the methodology and that the decision to release a particular micro level data set still depended to a substantial degree on judgement. The experiment made clear that the commonly used approach to releasing samples of anonymized census records has limitations for small countries. This is because a sample of census records from a very small country will be too small to be useful to a researcher when standard methods for creating samples of

census records are applied. This is a problem for countries with populations below one million and becomes particularly acute for countries with populations below 300,000 which would include many Caribbean countries.

20. The Population Affairs Officer highlighted the potential value of data archives in providing an important service to statistical offices and researchers. He said that data archives act as intermediaries between data producers and users, and they can specialize in disclosure control and managing microdata release to researchers. Further, he stressed the need for statistical offices to trust data archives to receive microdata and distribute it on their behalf. He suggested that collaborating with data archives could help countries mitigate their statistical offices' limited capacity to invest in microdata dissemination and disclosure control and proposed that countries deposit their microdata with the Derek Gordon Databank. He reasoned that a data archive that is located within the subregion can provide a more specialized service which meets the needs of Caribbean researchers, and the UWI databank has the potential to facilitate a major increase in microdata dissemination in the subregion. However, the Population Affairs Officer added that collaboration with IPUMS-International was equally valuable as the MPC is a world leader in dissemination of census microdata and provides a platform for worldwide dissemination at little or no cost to the NSO. Collaboration, not competition, is the key. He stated that ECLAC was already active in the area of census dissemination through REDATAM and stood prepared to support and advise on anonymization and disclosure control.

## **Discussion**

### **Technical/Statistical procedures**

21. The meeting discussed some of the technical procedures that could be used to anonymize data, including sampling, recoding, suppression and data swapping. Participants underscored the importance of balancing the protection of statistical confidentiality with the need to maximize data utility. The meeting also discussed the possibility that, for smaller countries, higher sampling fractions such as 20 per cent or higher, rather than 10 per cent, could be used to provide a larger sample of records. Increasing the sampling fraction might require greater suppression or recoding of other variables in order to achieve an acceptable level of disclosure risk.

### **Trainings and capacity building**

22. Meeting participants expressed interest in being more involved in the anonymization process and in receiving training in anonymization techniques. Some expressed further concern that their statisticians were not adequately trained in methods of statistical disclosure control.

### **Challenges to microdata dissemination**

23. Participants noted that the small populations of Caribbean Member States and the small pool of researchers presented major challenges to disseminating microdata. Additionally, because laws in the subregion protect the confidentiality of census records but do not explicitly authorize release of anonymized microdata. Although they do not prohibit release, statistical offices may be hesitant to disseminate microdata. Modernization of statistical legislation to provide a governing framework for microdata release would therefore help statistical offices to release more microdata, for example providing a legal basis for enforceable data access agreements.

### **Modes of data dissemination**

24. The meeting discussed the use of secure data laboratories in the Caribbean subregion. Mixed views were expressed regarding the viability of secure data laboratories as a means of disseminating census data. Such facilities do not tend to attract large number of users even in countries where one might expect there to be a large number of researchers interested in accessing detailed census microdata. Others felt that secure data laboratories remained a viable option. It was generally agreed that remote access to microdata was a particularly attractive option for small countries because remote access can be provided to full count microdata rather than samples of records.

25. In response to increasing requests for microdata, participants expressed a preference for licensed release of microdata rather than release through public use files which can be accessed by anyone. They cited concerns of data “falling into the wrong hands” and high risks of disclosure.

26. Participants also expressed their desire to continue to use REDATAM as a data dissemination tool. REDATAM provides remote access to census microdata for tabulation and calculation of basic indicators.

#### **Collaboration between data archives**

27. The representatives of the IPUMS-International project and the Derek Gordon Databank expressed their interest in collaborating with each other in order to make Caribbean census microdata more accessible to researchers.

#### **Commitment**

28. It was acknowledged that expansion of access to census microdata would not happen overnight but that progress would be achieved incrementally. Several participants emphasized that statistical offices needed to develop experience in the use of methods and tools for the anonymization of microdata. Data not only has to be made accessible but it has to be promoted among researchers to encourage its use.

29. Participants concluded that the meeting represented an important step toward institutionalising and increasing the dissemination of microdata and introduced some key methods and mechanisms for releasing microdata.

### **4. Data archives: Partners in expanding access to microdata**

#### **IPUMS-International: Revolution in census microdata access**

30. In his presentation, the Ambassador for IPUMS-International gave an overview of the data archive, which offers access to integrated and harmonized samples of census records. He discussed the IPUMS-International project, the organizations that participate, the users of the data, and the ways in which the IPUMS-International project could help expand access to Caribbean census microdata. He also conveyed his interest in working with ECLAC to ensure that the initiatives of statistical offices and other stakeholders were well coordinated. One of the reasons why coordination is important is that statistical offices can only release a limited number of microdata files from a given census because disclosure risks cumulate when multiple files are released.

31. The Ambassador for IPUMS-International explained some of the potential benefits to Caribbean statistical offices of collaborating with the project. IPUMS-International offers a system for managing access to “confidentialized” integrated microdata and metadata at no cost to NSOs. The Minnesota Population Center which runs the IPUMS-International project also assumes full responsibility for the storage and controlled dissemination of data from the archive. He suggested that it would be possible to develop a subregional web portal to promote access to Caribbean census microdata as well as national portals which would direct researchers from NSOs’ websites to the IPUMS International archive. There are plans to pilot remote access facilities in the near future. This will make it possible to provide access to detailed, full count microdata through software licensed by the Minnesota Population Center (MPC).

32. He summarized five key selling points of the IPUMS-International project. He stated that the project began in 1999, and was fully funded through to 2020. Second, he explained that his organization archived and documented census microdata for NSOs all over the world. Third, IPUMS-International presented a unique opportunity to integrate both microdata and metadata over time and across the world. Fourth, the organization disseminated licensed samples of records to researchers at no cost without ever releasing the original data. Finally, IPUMS-International was a means by which nations, regions, and the world could share integrated microdata, metadata and census boundary files.



33. NSOs have the option to send raw data to the MPC or anonymize the data themselves. The Ambassador for IPUMS-International assured the meeting that when NSOs provide full count microdata, only a very limited number of MPC staff can access it (he mentioned that he himself could not access it), and once a sample is drawn for release to researchers, the full data set is locked away. He also stressed that IPUMS-International had never had an incident related to misuse of data. The Ambassador for IPUMS-International further explained the “pick” extract system where users access precisely the variables, censuses, countries that they are interested in.

### **The Derek Gordon Databank**

34. The former manager of the Derek Gordon Databank, and ECLAC consultant, delivered a presentation on the work of the Derek Gordon Databank, part of the Sir Arthur Lewis Institute of Social and Economic Studies, at the Mona campus of the University of the West Indies. She explained that the databank was established in 1994 in response to the problem of limited availability of high quality microdata. The Derek Gordon Databank, which is located in Jamaica, is intended to serve the Caribbean subregion. Its primary purposes are to archive the data and provide access to researchers in order to facilitate and stimulate research.

35. The presenter described the procedures at the Derek Gordon Databank. When the databank receives a data set, some cleaning of the data may be carried out. It is then anonymized using basic techniques. Finally, the data are disseminated with databank staff preparing bespoke extracts for individual researchers. Where there is any doubt concerning the amount and level of detail being requested, the request is referred to the owner of the data to decide how to proceed with the request. She stated that the wishes of the owner of the data are always respected. The ECLAC consultant added that the databank also provides technical assistance and guidance to researchers upon request. She highlighted that the Derek Gordon Databank differs from IPUMS-International in that in addition to housing and population census data, it also stores and disseminates other surveys like literacy surveys and poverty assessment surveys.

36. She noted that the Derek Gordon Databank faced a number of challenges. The databank had a good collection of data sets from Jamaica but had faced reluctance from statistical offices outside Jamaica to deposit data. Consequently the databank primarily served researchers interested in Jamaica. She suggested that the databank suffered from a lack of visibility, and the absence of a strongly developed culture of data sharing, social research and evidence based policymaking in the Caribbean. She also suggested that across the subregion in general there was an overly conservative approach to microdata dissemination which was not based on a realistic assessment of the risks of disclosure. She confirmed that the databank has been disseminating microdata for around 20 years without any known breach of confidentiality.

## **Discussion**

### **Depositing data at IPUMS-International and the Derek Gordon Databank**

37. Participants sought clarification on the procedures for NSOs to share data with the data archives and inquired about the confidentiality agreements between NSOs and the data archives.

38. The representative of the IPUMS-International project stated that the first step to collaborate with the project was the signing of a memorandum of understanding. He explained that countries could simply gather their documentation and microdata and submit these data to IPUMS-International, which would then “confidentialize” the data as necessary in preparation for dissemination. He explained that countries also had the option to draw their own samples and anonymize the data themselves before submitting it to IPUMS-International. He informed the meeting that the integration process could take three to five years to complete once the data had been received.

39. The representative for the Derek Gordon Databank explained that terms of use are negotiated between the Derek Gordon Databank and the respective NSOs. These terms of use generally specify conditions that the researcher must meet in order to access the data, restrictions on their use of the data, and they also stipulate how authors should acknowledge their use of the data in publications. At the request of an NSO, the databank would also resubmit the cleaned and anonymized data to the NSO to ensure that all parties are satisfied with the data before they are disseminated. She reassured the meeting that the databank did not make the raw data available to anyone.

#### **Technical assistance and workshops for NSOs and researchers in the subregion**

40. The Ambassador for IPUMS-International acknowledged the important role that the Derek Gordon Databank could play in offering both survey data and a more personalized service, sometimes including technical advice to researchers. He noted that while IPUMS-International did not typically clean census data, if necessary, they could provide an expert who would work with the NSO to clean the data. He suggested that, while training workshops are normally reserved for larger countries, IPUMS-International may consider doing a training workshop in the Caribbean subregion based on the capacity and needs of the NSOs. IPUMS-International may also provide technical assistance to NSOs if their data prove to be particularly problematic.

#### **Collaboration between the data archives**

41. Each data archive acknowledged that the other had unique services to offer the Caribbean subregion. As such, they committed to collaborating with each other to provide the best service possible in terms of providing researchers with access to anonymized microdata from the subregion.

#### **Increasing visibility of data from the Caribbean subregion**

42. The IPUMS-International representative suggested that since the Derek Gordon Databank was already using the Data Documentation Initiative (DDI) metadata standards, the DDI metadata could be provided to the World Bank to post on their Central Microdata Catalog website in order to generate new users for the Derek Gordon Databank. He described how providing the metadata for data sets in the IPUMS-International archive had increased visibility and was now generating 15 per cent to 20 per cent of the new users each year.

### **5. Closing remarks**

43. Closing remarks were delivered by the Coordinator of the Statistics and Social Development Unit on behalf of the Director of the ECLAC subregional headquarters for the Caribbean. He expressed appreciation for the contributions made and thanked everyone for participating in the meeting. He restated the goal to have more microdata made available to researchers. He expressed his support for all parties to collaborate, not compete, in order to achieve better data dissemination to support research and policymaking in the subregion. He assured the meeting that ECLAC intended to work with countries to overcome the cultural, technical, legislative, and administrative obstacles to the release of microdata in the subregion. He also assured the meeting that the finalized paper would be distributed widely.

Annex I**LIST OF PARTICIPANTS**

Halim Brizan, Director, Central Statistical Office, Ministry of Finance, Planning, Economy, Energy, and Cooperatives, Grenada. E-mail: hmbrizan@gmail.com

Tiemonne Charles, Statistician/Software Engineer, Central Statistical Office, Ministry of Finance, Planning, Economy, Energy, and Cooperatives, Grenada. E-mail: tiemonne@gmail.com

Helen Kristin Fox, Consultant and former Manager, Derek Gordon Databank, University of the West Indies, Mona campus, Jamaica. E-mail: kazfox3@yahoo.com

Boyd Goodin, Director, Information and Technology Division, Statistical Institute of Jamaica. E-mail: hes@statinja.gov.jm

Richard Leach, Derek Gordon Databank, University of the West Indies, Jamaica. E-mail: richard.leach@uwimona.edu.jm

Robert McCaa, Research Professor/Ambassador, IPUMS-International, Minnesota Population Center, University of Minnesota, United States of America. E-mail: rmccaa@umn.edu

Janet Martin, Acting Director, Censuses, Demographic and Social Statistics Division, Statistical Institute of Jamaica. E-mail: sesu@statinja.gov.jm

Sandra Paredez, Technical Advisor on Censuses, United Nations Population Fund (UNFPA), Jamaica. E-mail: paredez@unfpa.org

Hope Perkins, Research Officer, Statistical Institute of Jamaica.

Gatlin Roberts, Chief Statistician, Statistical Office, Ministry of Finance and Economic Planning, Saint Vincent and the Grenadines. E-mail: svg.stats@mail.gov.vc

Marva Williams, Assistant Director of Statistics, Central Statistical Office, Ministry of Planning and Sustainable Development, Trinidad and Tobago. E-mail: marvadeewill@gmail.com

**Economic Commission for Latin America and the Caribbean****Subregional headquarters for the Caribbean**

Abdullahi Abdulkadri, Coordinator, Statistics and Social Development Unit. E-mail: abdullahi.abdulkaadri@eclac.org

Francis Jones, Populations Affairs Officer, Statistics and Social Development Unit. E-mail: francis.jones@eclac.org

Dale Alexander, Computer Information Assistant, Programme Support Unit. E-mail : dale.alexander@eclac.org

Candice Gonzales, Research Assistant, Statistics and Social Development Unit. E-mail: candice.gonzales@eclac.org

Luanne Serieux-Lubin, Research Assistant, Statistics and Social Development Unit. E-mail: luanne.serieux@eclac.org

Tanisha Ash, Statistical Assistant, Statistics and Social Development Unit. E-mail: tanisha.ash@eclac.org