National Practices on Household Survey Microdata Dissemination

April 2022

Prepared by Masters Students at:

LSE Department of International Development

Prepared for:

Statistics Division

March 2022

April 2022 Revision

The report has been created by Christine Imali, Emily Campbell Harry, Lindsay Caird, Mahima Chaudhari who a currently MSc students in the Department of International Development at the London School of Economics and Political Science.

Disclaimer

The views represented in this report are those of the authors. This report does not represent the views of the London School of Economics and Political Science nor the United Nations Statistical Division or the Inter-Secretariat Working Group on Household Surveys.

This report has been designed using images from Flaticon.com and Freevectormaps.com

Table of Contents

Acknowledgements

National Practices on Household Survey Microdata Dissemination is commissioned by the United Nations Statistics Division, Inter-Secretariat Working Group on Household Surveys (ISWGHS).

We would like to acknowledge the following people for their encouragement and support throughout the project:

Eric Swanson, Open Data Watch
Haoyi Chen, United Nations Statistical Division
Nemi Okujagu, United Nations Statistical Division
Tiziana Leone, London School of Economics and Political Science (LSE)

We would also like to thank the respondents from Costa Rica, State of Palestine, United Republic of Tanzania, and Vietnam for taking the time to share with us their experiences and knowledge.

Acronyms

ASEAN - the Association of Southeast Asian Nations
CCSA - Committee for the Coordination of Statistical Activities
DCMI - Dublin Core Metadata Initiative
DDI - Data Documentation Initiative
FAO - Food and Agriculture Organization of the United Nations
FTP - File Transfer Protocol
GHDx - Global Health Data Exchange
GSO - General Statistics Office of Vietnam
IHSN - International Household Survey Network
ILO - International Labour Organization
IMF - International Monetary Fund
INEC - Instituto Nacional De Estadística Y Censo (National Institution of Statistics & Census of Costa Rica)
ISWGHS - Inter-Secretariat Working Group on Household Surveys
JSON - JavaScript Object Notation
LMIC - Low- and Middle-Income Countries
LSE - London School of Economics and Political Science
NADA - National Data Archive
NBS - National Bureau of Statistics
NSDS - National Strategy for the Development of Statistics
NSO - National Statistics Office
NSS - National Statistics System
ODB - Open Data Barometer
ODC- International Open Data Charter
ODIN - Open Data Inventory
OECD - Organization for Economic Co-operation and Development
PCSB - Palestine Central Bureau of Statistics
PUFs - Public Use Files
SDC - Statistical Disclosure Control
SRC- Scientific Research Center
SWOT - Strengths, Weaknesses, Opportunities, Threats
TNADA - Tanzania National Data Archive
TSED - Tanzania Socio-Economic Database
UN - United Nations
UNDESA - United Nations Department of Economics & Social Affairs
UNFPOS - United Nations Fundamental Principles of Official Statistics
UNGA - United Nations General Assembly
UNSD - United Nations Statistical Division
VSDS - Vietnam Statistical Development Strategy 2021 - 2030, vision to 2045
WB - World Bank

Definitions

Anonymisation - The process of rendering data into a form that does not identify individuals and where identification is not likely to take place (Information Commissioner's Office, 2012).

Direct Identifiers - Refers to data values that identifies an individual in a data set, for example names, addresses, telephone numbers, and unique numbers, in addition to other identifiable information such as social security numbers (Innovative Routines International, 2022).

Household Surveys - Questionnaires given to a sample of households in a population, providing considerable information about the respondent/individuals (Wolff, 2015).

Indirect Identifiers - Refers to attributes that may be true about an individual but are not necessarily unique to an individual and could be used to attempt to identify respondents. These include detailed geographical information, composition of household by age, sex, and detailed information on professional status (Innovative Routines International, 2022).

Metadata - A description that summarizes basic information about data that makes it easier to find, use, and re-use data. It is often called data about data (National Information Standards Organization, 2004).

Microdata - Electronic data files containing the information about each unit of observation in household surveys, censuses, or administrative records (Dupriez & Boyko, 2010).

National Statistics Office (NSO) - The main producer of official statistics in a country that may also be responsible for coordinating all activities related to the production and dissemination of official statistics (UNDESA, 2021).

Executive Summary

The report on the National Practices on Household Survey Microdata Dissemination, prepared by post-graduate student consultants at the London School of Economics and Political Science (LSE) in collaboration with the Inter-Secretariat Working Group on Household Surveys (ISWGHS) of the United Nations Statistics Division (UNSD), examines the practices and challenges of National Statistics Offices (NSOs) disseminating household survey microdata.

The report was produced in response to the increasing value of household surveys in important decision-making processes and research purposes, combined with the evolving nature of microdata management and international dissemination recommendations. It was further motivated by the lack of existing research on the microdata dissemination practices of NSOs in low- and middle- income countries (LMICs), where resource and financial constraints, limited technical and training capacities as well as unclear privacy protection rules create additional barriers for data producers. Thus, this report addresses the following research questions:

- What are the prevailing practices in microdata dissemination versus the recommended practices?
- What are the enabling legal environments, including national statistical law and data privacy law?
- What considerations are made on data anonymisation and protection?
- Microdata accessibility: In what format are data disseminated and who has access to them?

To answer these questions, the report concentrates on four case studies, Costa Rica, State of Palestine, the United Republic of Tanzania and Vietnam, and presents the insights of local microdata dissemination experts through country profiles.

Primary qualitative data was collected through desk research as well as interviews and a focus group with NSO statisticians from the four countries indicated above. A total of six key informants were interviewed. A comparative case study approach was used to analyze the transcribed interviews, offering an in-depth contextual consideration of each country's practices whilst enabling cross-case themes to be explored.

An assessment of the NSOs current practices was undertaken by drawing comparisons with the UN's recommended best practices and other guides, such as those produced by the International Open Data Charter (ODC), Open Data Inventory (ODIN), International Household Survey Network (IHSN), Organization for Economic Co-operation & Development (OECD), and UN General Assembly (UNGA). Through this, the report identifies challenges faced by countries that hamper them from achieving efficient data dissemination practices.

Key Findings

Legal Environment

Legislative frameworks are available in each of the four countries, with policies that guide the collection and dissemination of national statistics and microdata. These statistical laws and policies have been amended over the years to reflect the UN Fundamental Principles for Official Statistics, although the degree of execution varies across countries.

Data Security and Anonymisation

Limited capacity and resources prevent staff training on microdata anonymization and dissemination. A challenge further reflected by the NSOs limited capacity to monitor microdata use once it is disseminated.

Access to Microdata

Diverse ability to access microdata based on country. Some countries lack online microdata repositories, others have repositories that are not fully functional. Lack of repositories or limited functioning repositories result in inefficient and time-consuming microdata dissemination.

Key Recommendations

Training

Collaboration is required between development partners and countries to enhance training and benchmarking for NSOs, as well as the sharing of knowledge on microdata anonymisation, packaging, dissemination and any additional technical techniques.

Consistent Guidelines

Interagency coordination to produce an abridged version of microdata dissemination guidelines that are universally consistent in regard to the security, access, legal frameworks, and the technical aspect of dissemination to provide a streamlined and cohesive approach for countries.

Funding

Countries require low-cost options for data portals. In addition, highlight the cost-effectiveness of abiding by the best practices guidelines in a tangible format to encourage compliance and to encourace governments and other development partners to invest in NSOs.

Data Security

Standardization of dissemination guidelines and best practice recommendations regarding how to monitor data users' output to promote confidentiality and appropriate use of provided data.

Technical Support

The development or increasing the ability to access user-friendly data portals that are embedded with microdata request forms and metadata download functions to guarantee ease of access, as is the case with international organization's repositories

Introduction

Microdata Dissemination in the Era of Open Data

Parallel to the accumulating power of computers and software, the demand for microdata has exploded in recent decades (Dupriez & Boyko, 2010). Obtained from sample surveys, censuses and administrative records, microdata underlies published statistics, containing information on individual respondents, households, and business entities (ibid; UN, 2021). As such, the management of microdata is a key responsibility of national statistics agencies who face conflicting objectives when disseminating microdata to the public. On one hand, National Statistics Offices (NSOs) aim to release useful data files in support of diverse statistical analysis, innovative research, and essential decision- making practices; on the other, it's required to safeguard the anonymity of informants' identities (Gomatam et al, 2005). Hence, data is typically 'sanitized' before being released, with direct and in-direct identifiers being removed, re-coded or swapped, and noise added to prevent re-identification (Reiter, 2012; Willenborg & de Waal, 2001). Unfortunately, such statistical disclosure protection techniques reduce the accuracy of the data, yet the complex process is necessary to uphold respondents' trust and privacy legislation through confidentiality protection (Duncan et al, 2001). In this light, disseminating microdata is not only a technical and organizational obstacle for NSOs, but a legal and ethical challenge (Dupriez & Boyko, 2010). Consequently, the role of government in promoting the broad use of microdata has evolved, becoming paramount in facilitating microdata accessibility, data anonymisation, and coinciding legislation, reflective of international guidelines that were developed to address such issues.

Figure 1
Organizational Best Practices edited for length (CCSA, 2014)

Openness	Access must be provided on equal terms for the international research community at the lowest possible cost, preferably at no more than the marginal cost of dissemination.
Transparency	Detailed metadata must be provided, and the specifications of conditions attached to the use of the data should be internationally available in a transparent way, ideally through the internet.
Legal conformity and protection of privacy	National laws and international agreements, as they pertain to the protection of privacy, directly affect data access and sharing practices. These must be taken into account in the formulation of data access arrangements. Unless formal consent has been provided by the respondent data can *only* be disseminated after being properly treated to ensure the risk of disclosure is minimal
Protection of intellectual property	Data access arrangements must consider the applicability of copyright or of other intellectual property laws that may be relevant.
Interoperability	Technological and semantic interoperability is key in enabling and promoting international and interdisciplinary access to and use of research data. Access arrangements should pay attention to the relevant international data documentation standards.
Quality	The value and utility of data depend, to a large extent, on the quality of the data itself. Data managers, and data collection organizations, should pay particular attention to ensuring compliance with explicit quality standards
Security	Specific attention should be devoted to supporting the use of techniques and instruments to guarantee the integrity and security of data
Accountability	The performance of data access arrangements should be subject to periodic evaluation by user groups, responsible institutions and funding agencies.

Since the turn of the century, consensus on statistical practices regarding the dissemination of microdata has been assembled into numerous best practice guidelines, handbooks, and software tools by a variety of organizations, including the United Nations Statistics Division (UNSD), the United Nations (UN) Economic Commission for

Europe (2003; 2007 & 2009) and the Organisation of Economic Co-operation and Development (2007). These key principles, international technical standards and recommended practices have largely centered on the UN (2014) Fundamental Principles of Official Statistics (UNFPOS), specifically Principle One and Six. Principle One provides the basis for the dissemination of microdata, stating, in part, “official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens’ entitlement to public in formation” (ibid). Taking this ‘open-data’ proposition into consideration, an online repository is arguably the most important tool used by an NSO to transparently distribute open and free datasets, including disseminated microdata and metadata (UN, 2021).

Figure 2
Organizational Best Practices edited for length (CCSA, 2014)

Documentation	Data Documentation Initiative (DDI) by the DDI Alliance is the most commonly used standard for microdata. Adopting the DDI as a UN recommended standard for microdata documentation would bring the UN institutions in line with current best practice
Cataloguing	To make data discoverable detailed metadata should be published in on-line searchable catalogs. Compliance with the DDI standard makes it considerably easier. Open source DDI compliant cataloging applications already exist such as the open source National Data Archive (NADA) or DataVerse
Anonymization	No international standards exist for the implementation of anonymization and institutional and national practices are typically not documented or shared. However, the methods used for anonymizing microdata are very contextual to the type of data being anonymized, and disclosing detailed information on the methods may provide useful information to those trying to defeat the protections.
Dissemination	It is important to understand that not all data are the same. As such, dissemination policies that are clear but also flexible enough to cover the full range of issues such as data ownership, legal and ethical responsibilities and sensitivity of data need to be developed.
Preservation	Procedures and infrastructures must be put in place to protect data against hardware and software obsolescence (regular migration of datasets to new media and formats), system failures, human errors and other hazards.

Microdata is further governed by Principle Six, stating that “individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes” (UN, 2014). This principle underlies microdata dissemination policies for NSOs and international organisations, in addition to the organisational best practices summarized in Figure 1 and the technical standards in Figure 2. Even with the availability of these guidelines and tools, the ‘open-data’ movement remains under-developed in many low- and middle-income countries (LMICs). Here, NSOs often face additional barriers such as resource and financial constraints, limited technical and training capacities as well as unclear privacy protection rules, that exacerbate the multifaceted tradeoff between disclosure risk and data usefulness (Dupriez & Boyko, 2010; UNSD, 2014).

Despite the need for research to understand the barriers and benefits regarding microdata dissemination across different geographical and economic contexts, existing research is lacking. Instead, research concentrates on international organisation practices and Western legislation (Ruggles, 2014; UNSD, 2014).

Therefore, this report seeks to confront this research gap by investigating microdata dissemination practices of NSOs in different contexts to address the following questions:

- What are the prevailing practices in microdata dissemination versus the recommended practices?
- What are the enabling legal environments, including national statistical law and data privacy law?
- What considerations are made on data anonymisation and protection?
- Microdata accessibility: In what format are data disseminated and who has access to them?

To explore these questions, the report will first detail the methodology, before presenting the case studies of Costa Rica, the State of Palestine (referred to as Palestine), the United Republic of Tanzania (referred to as Tanzania), and Vietnam, drawing on primary qualitative research collected through key informant interviews with statisticians from each country's NSO (Table 1). Thereafter, in the discussion section, countries current practices will be compared to recommended practices, identifying prevailing challenges and gaps in LMICs existing microdata dissemination. The report concludes with recommendations on various approaches and tools to support countries working at different microdata dissemination capacities.

Table 1
Case Study NSOs

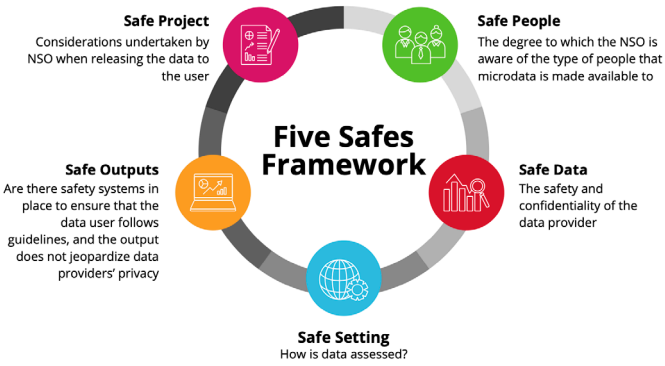
COUNTRY	NAME OF NSO	NSO LOGO
Costa Rica	National Institute Of Statistics And Censuses	
State of Palestine	Palestinian Central Bureau Of Statistics	
The United Republic Of Tanzania	Tanzania National Bureau Of Statistics	
Vietnam	General Statistics Office Of Viet Nam	

Methodology

Research Design

Primary qualitative data was collected for this project through key informant interviews and a focus group discussion with staff from the NSOs of Costa Rica, Tanzania, Vietnam, and Palestine. These respondents were chosen in consultation with the Inter-Secretariat Working Group on Household Surveys (ISWGHS), selecting countries on the premise of their different geographical locations and the representation of varied statistical development. Prior to the key informant discussions, an interview guide (Appendix D) was developed, with questions organised under the 'Five Safes' framework (Figure 3). This framework was used to elicit detailed explanations of data dissemination practices, considering the use of confidential data and decision-making processes of the NSO. Between the 19th of January and 28th January 2022, three key informant interviews and one focus group were conducted with statisticians from the four NSO's mentioned previously. The discussions took place over a private Microsoft Teams meeting, each lasting 45 to 120 minutes. One of the interviews was facilitated by a translator, whilst the remaining three were conducted in English, accounting for the variation in interview duration. Nevertheless, after ascertaining permission from the respondents, all interviews were recorded and transcribed.

Figure 3
The Five Safes Interview Framework



Sample

To identify respondents for this report, purposeful sampling was used in conjunction with snowball sampling. Participants were suggested in accordance with their knowledge and expertise on the phenomenon of interest, maximizing efficiency and validity of contributions (Johnson, 2014; Palinkas et al, 2015). Potential respondents from the associated NSOs were limited to statisticians' knowledgeable about microdata dissemination practices and data privacy legislation at the national level. Initially, respondents were recruited by employing existing relationships between the UNSD and NSOs, although these primary contacts subsequently referred to statisticians affiliated with the relevant criteria. It should be noted that although snowball sampling is often advantageous in increasing the likelihood of participation response, the original sample size was decreased from five countries to four after the fifth NSO failed to uphold an original interview commitment (Robinson, 2014). Hence, a total of six informants were interviewed across the remaining four countries.

Analytical Approach
Transcribed interview responses were analysed using a comparative case study approach. As a study design, case studies are often adopted to advance understanding of a particular issue, gaining insight from a variety of data sources such as interviews, observations, documents, and databases (Elkatawneh, 2016; Hyett et al, 2014). This approach was selected because it allows an in-depth consideration of each case, regarding institutional and political contextual factors, and case specific themes whilst enabling cross-case analysis and a proposal of lessons learned (Patton, 2015; Stake, 1998).

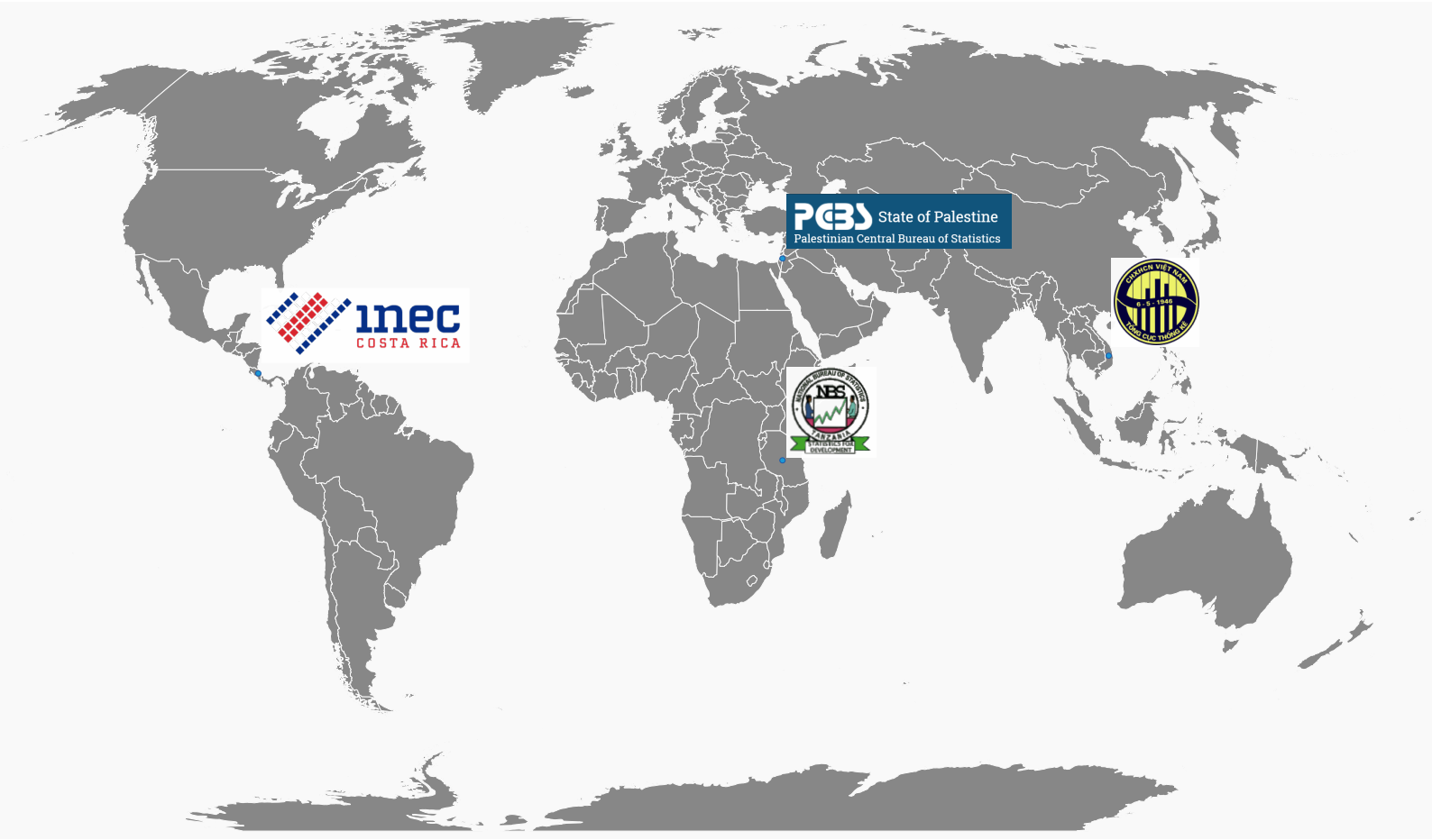
Ethics
Prior to conducting the interviews, ethical approval was obtained from the London School of Economics and Political Science (LSE) internal review board. In compliance with the LSE's guidelines and ethical codes, written informed consent was obtained from each respondent during the interview process. Furthermore, data was stored on a password protected drive, ensuring confidentiality throughout.

Limitations
Considering time constraints and the scope of this report, the number of countries was limited to four. This limits the representation of statistical development and generalisation of the findings and coinciding recommendations. However, in-depth interviews have provided detailed, comparable data to present robust recommendations based upon analysis and discussions.

The semi-structured interviews encouraged rich communication between interviewers and interviewees, facilitated by open-ended questions on the subject matter. Due to the sensitivity of recruitment through snowball sampling and the nature of this report, no pilot studies could be conducted prior to the interviews. To overcome this limitation, several drafts of the interview guide were produced covering the five core, comprehensive dimensions of data dissemination before being confirmed with the client. Additionally, the semi-structured format allowed follow up questions to be asked throughout the interview and respondents agreed to answer further questions through email correspondents where necessary.

This qualitative study relies on the impressions, opinions, and views of respondents; therefore, findings are interpreted in accordance with participants' subjective understanding of the topic. Consequently, the reliability, validity, and generalisability of qualitative results is often scrutinised (Nobel & Smith, 2015). Nevertheless, the predetermined expertise of the participants enhances the reliability of the findings, and the approach offers constructive insight into the complexities of data privacy.

Country Profiles



Costa Rica

Data Dissemination Landscape

The National Statistics System (NSS) in Costa Rica is characterized by a decentralized structure, a well-developed legal framework, an accessible online repository with tiered microdata disclosure, and a respect for confidentiality, making Costa Rica largely compliant with the UNSD microdata dissemination best practices (CCSA, 2014). Costa Rica is an administratively decentralized unitary constitutional republic, nevertheless, the Instituto Nacional De Estadística Y Censos (National Institute of Statistics and Census of Costa Rica), or INEC, represents the central agency within the NSS. Since its establishment in 1998 with Law No. 7839, the INEC has been the main coordinator of official statistical activities and, according to Article 4, authorises the delivery of microdata in Costa Rica (OECD, 2019). As such, considerable progress has been made in the quality of statistical practices, namely the INEC increasingly disseminates statistical information on its online repository in accordance with international guidance. Of the eighteen staff within the INEC, two statisticians work directly with the dissemination of microdata, however no formal training is given to the staff on dissemination practices when joining the department. Despite the update in statistical and data laws, policy for the continuous vocational training for INEC staff remains absent, although guidelines are accessible online for all personnel[1].

Legal Framework

Over the past decade, Costa Rica's legal framework surrounding microdata dissemination has been updated. To further the principles enacted in Article 4 of the 1998 Law No. 7839, the INEC adopted a new Dissemination Policy for Statistics and Delivery of Microdata in December 2014; a policy that is being extended to the NSS (ibid; National Institute of Statistics and Censuses, 2014). Consistent with the UN Fundamental Principles for Official Statistics, the policy requires data to be collected and produced following scientific norms on sources, methods, standards, and procedures of statistics, maintaining confidentiality of individual information when disseminating microdata and metadata for free (OECD, 2019). The issue of data anonymity is further supported by Law No. 9694, a law adopted in May 2019 to amend the initial 1998 Law No. 7839, putting Costa Rica in line with the Recommendation of the Council on Good Statistical Practice (ibid). The Law No. 9694 strengthens data privacy laws by ensuring protection of personal data and regulating pre-release access arrangements (ibid). Moreover, the new policies recognise the INEC as an autonomous institution, sustaining adequate financing for the INEC, improving the access to administrative sources, and providing a legal mandate for the coordination of official statistics and microdata dissemination.

[1] Informed in the interview that these guidelines were accessible online but cannot be located through desk research

Microdata Accessibility

Microdata is accessible through Costa Rica's National Archive (ANDA for its Spanish Acronym) and an online repository managed by the INEC. The ANDA is in Curridabat, San José, less than one kilometre away from the INEC headquarters (Archivo Nacional, 2022). Here, all documented statistical operations, disseminated databases, and anonymised microdata files are available for the user. Furthermore, the INEC has designed an easy-to-use web platform for the dissemination of microdata and other official statistics produced in the NSS. This online repository exclusively stores datasets produced by the INEC from 2001-2022. Completed metadata is available to download in the format DDI/XML & JSON, whilst microdata is available to download locally through an SPSS Statistical Data File or SAP format.

Under the strict protocols of the Dissemination Policy for Statistics and Delivery of Microdata, the INEC has established a clear procedure for microdata dissemination and its accessibility for external users. Differing in degrees of sensitivity, three types of microdata files have been generated: (i) Public Use Files (PUFs) (ii) licensed files (iii) non-accessible files. The PUFs are accessible to any citizen and can be downloaded once a profile has been set up using a valid email address. Although users must provide a general description of purpose and data use, requests are not authenticated by the INEC and the user's output is not reviewed. Therefore, these microdata files are open without restrictions. The anonymisation of microdata is less rigorous for licensed files, where indirect identification may be possible despite the elevated levels of disaggregation. For this reason, licensed files are restricted to recognised institutions and verified academic researchers. These licenses are free of charge. When requesting access, researchers must disclose in detail, the justification and methodology of their proposed study and are expected to share the results with the INEC upon publication. Additionally, legal representatives from the organization or research institution are required to sign the Contract for Access to Microdata, agreeing to the Law's strict terms of confidentiality (National Institute of Statistics and Censuses, 2014). When it is not possible to fully anonymise datasets, files are classified as non-accessible. This microdata is particularly confidential and sensitive as the data contains easily identifiable information, such as farm sizes and number of company employees. Therefore, access is restricted to authorized INEC personnel. Information on the type of survey and its assigned tier is presented in Table 2. Abiding by these compliances, microdata is further disseminated through other international organization repositories. Microdata and metadata are saved locally when produced in collaboration with the INEC and are accessible after logging-in to the repository. Where microdata is not stored locally, links to the INEC repository are provided, however these links do not always work. Details on international organisation repositories are provided in Appendix F.

Table 2

Type of Survey and Access

Type of Survey	Type of Microdata File
National Household Survey	Public
National Agriculture Survey	Non-Accessible
Continuous Survey of Employment	Public
Survey of Women, Children and Adolescents	Public
National Culture Survey	Public
National Disability Survey	Public
National Survey of Producing Households/National Survey of Household Microenterprises	Public
National Survey of Time Use	Public
National Survey of Jobs in Private Establishments, Agricultural Sector, Manufacturing and Construction	Non-Accessible
National Survey of Household Income and Expenditure	Public
Labor Force Survey	Public & Licensed (less anonymized, more raw data)

Data Anonymity and Protection

In accordance with the UNFPOS (2014), the current laws and practices of the INEC offer protection of statistical safety and confidentiality. The INEC eliminates the most sensitive, direct variables from data sets, whilst indirect variables are collapsed, aggregated, and recoded to hide identifiable information. As mentioned, this process of anonymous animation is less rigorous for the licensed datasets. Although extensive effort is made to eliminate and hide variables, indirect variables risk reidentification, for example through geospatial identification, hence microdata files are released under a safeguarding agreement at a lower level of anonymity. The INEC notes that improving this anonymisation process for licensed databases is a current target for the NSS.

To protect the security and integrity of this online archive, a data security system for the internal archive has been implemented following the 2014 Dissemination Policy for Statistics and Delivery of Microdata. Regarding non-accessible datasets, an access code is required which is restricted to specific personnel within the INEC. Meanwhile, for license holders, an exclusive FTP is sent to the external researcher with an encrypted, temporary password that gives the authorized individual one week to download the files. Although the license agreement contains specific security protocols, it should be noted that compliance is not audited due to the INEC's lack of verifying capacity, thus FTP downloads could be passed onto additional parties without INEC authorisation. This is a security challenge that the INEC is seeking to improve. Nevertheless, new legislation is strict, imposing extensive legal ramifications, including fines, imprisonment, or sanctions, for any public or employee infringement of microdata confidentiality (ibid; OECD, 2019).

This is partly in response to a previous incident concerning the National Agriculture Survey, when identifiable information was shared with a sponsor, the Central Bank, who subsequently contacted data providers. Due to discrepancies between the Central Bank and INEC legislations, no further sanctions were taken but this has since been amended through the new 2014 data privacy law. INEC notes that they have learned from this event and under the new disclosure policy, there has been no breach of the security system or instances of nonconformity to privacy laws at the time of authoring this report.

State of Palestine

Data Dissemination Landscape

Established in 1993, the Palestinian Central Bureau of Statistics (PCBS) is a centralized, independent, and transparent body that produces reliable, country-wide official statistics, which are subsequently used by the Palestinian government to formulate policies. The PCBS obtained its legal status as the official data collection body after the Palestinian Statistics General Law was passed in 2000. All PCBS published data is stored within a secured online repository in English and Arabic, making it easily accessible and comprehensive to users fluent in these languages.

The PCBS follows a nationally approved Official Statistics Code of Best Practice in compliance with the UNSD microdata dissemination best practices (CCSA, 2014). This code strengthens the legal, administrative, and financial functioning of the repository by outlining tasks carried out by its members (PCBS, 2006). There are a total of 258 members in the statistical division. The organizational structure is headed by a President, who is aided by two assistants - one for complementary purposes and another for statistical programs. There are a total of thirteen directorates with each consisting of three to five departments (PCBS, 2022). Although PCBS's wide data dissemination processes involve a sizable proportion of staff members, no training is provided specifically for dissemination of statistics.

Nevertheless, information on the process of microdata dissemination, including detailed descriptions on the methodology, sampling, and data collection, is available in the user's guide located in the data catalogue system. Further, Palestine has made consistent efforts to meet the international standards of microdata dissemination by setting up a specialized advisory council, specifically for this purpose. It includes government and non-governmental representatives, private sector, and academics but is neither part of PCBS nor included in any technical or administrative activities.

Legal Framework

In 1995, PCBS adopted its first statistical master plan that has since been renewed every four years. Each new plan is updated by conducting a SWOT (Strengths, Weakness, Opportunities, Treatment) analysis of the previous plan, making necessary changes to employ the best statistical practices and meet the evolving needs of national and international users (PCBS, 2009). The General Statistics Law No. 4 of 2000 defined the purpose of PCBS as a unified system creating comprehensive data to diagnose problems and evaluate progress (PCBS, 2009). With its introduction, PCBS witnessed improvement in the confidentiality of individual records as it required all employees to sign an affidavit certifying the privacy of data providers. Additionally, it adopted a policy requiring the removal of direct and indirect identifiers from raw data, publishing it as aggregate tables (PCBS, 2021). The law attempted to minimize the cost of data collected by preventing its recollection and duplication, achieved by ensuring harmony between the PCBS and government institutions.

The Fundamental Principles of Official Statistics that were adopted by Palestine in 2005 revised the guidelines and principles laid in the preamble for data dissemination. These stressed the responsibility of statisticians to serve the government and public by providing economic, demographic, social, and environmental data. In response to this, the PCBS subscribed to the General Data Dissemination System (GDDS) of the International Monetary Fund (IMF) in 2006. The PCBS dissemination policy focuses on accessibility, relevance, quality, and coordination, areas that have been highlighted in the UNSD microdata dissemination best practices (CCSA, 2014).

Microdata Accessibility

The adoption of the Creative Commons Attribution License in 2005, specified that PCBS views statistical data as a public good and mostly offers it free of charge to beneficiaries, without prioritizing or discriminating against any user. It aims to maximize reach to beneficiaries and minimize their efforts in accessing data through effective dissemination. Although there is no national repository, over time Palestine has transitioned from old methods of dissemination, such as paper and CDs, to different forms of online dissemination (PCBS, 2014). Whilst this includes various data visualization techniques, data is primarily disseminated online through the official PCBS website which contains disseminated microdata for 93% of all surveys conducted in the last 10 years[2]. Along with datasets, the website also provides a user guide, questionnaire and methodology used to carry out the survey. Here, Palestine's metadata is formatted using the international standards of Dublin Core Metadata Initiative (DCMI) and DDI, accessed through the Nesstar Publisher application.

The procedure to access microdata begins with the user completing an online form at the Division of User Services. In this form, the requesting persons must describe the data's intended use and the research being conducted. Most reasons are acceptable as no specific qualifications are required and there have been no instances where PCBS has denied access. Data requests are predominantly received from national and international organizations or university students conducting research. Students need to provide a certificate of registration to prove the data is for academic purposes. After submitting, PCBS takes up to three working days to process and approve the user's request. Once accepted, PCBS conducts a preliminary check of the user to make sure that the request is legitimate. Further, the applicant may be contacted to sign an agreement known as the License Agreement for the use of Microdata, stating that data will only be provided for scientific purposes. Researchers can also work in the Scientific Research Center (SRC) to obtain output, which is provided by PCBS after removing identifiers and other confidential data.

All PUFs are disseminated on the PCBS website, and some surveys are also available for download on the websites of international organisations like the International Labour Organization, Global Health Data Exchange, and the World Bank (see Appendix F). However, most of these websites require a mandatory log-in to access data, a request that can take up to three business days to be approved. Furthermore, because many of these international organisations do not save microdata locally on their websites, users are directed to the main PCBS webpage. Once the user has fulfilled all requirements, the data can be downloaded.

[2] Unable to independently verify this through desk research.

Data Anonymity and Protection

The PCBS respects the confidentiality of users and data providers. Data are anonymized by removing certain direct and indirect identification variables. These include the first and last name, name of the institution, addresses of residence of individual, family, and building number. Some indirect identifiers, such as rare scientific majors or institution-specific activities, are deleted and variables are recoded. This is done by changing the serial number of the sampling unit, using linear transformation of the serial number.

A security system regulates the certification and privacy of microdata. It ensures that users are unable to access microdata before signing the license agreement and has thus far never faced a breach. However, PCBS has policies set up to tackle any dispute or breach of laws. If the user violates the conditions outlined in the agreement, both parties seek a tripartite tribunal headed by a mutually accepted statistician.

United Republic of Tanzania

Data Dissemination Landscape

The National Bureau of Statistics (NBS) was founded under the 1997 Executive Agencies Act on March 26th, 1999. The ensuing 2015 Statistics Act established NBS as an autonomous public office, mandated to provide official statistics to the government and the public. It is the depository for all national statistics conducted in Tanzania. The NBS website lists the surveys carried out in Tanzania under the publications tab, with household surveys saved under poverty indicator statistics and health statistics. Here, the website has links to various databases where pdf reports and aggregated data compilations in the form of charts and graphs for household surveys are available. In addition, the NBS website is linked to the Tanzania National Data Archive (TNADA) database that is the country's national microdata repository. TNADA is hosted externally by the Tanzania e-Government Agency which is responsible for hosting all government portals. Overall, the distribution and dissemination of official statistics is coordinated by the IT and Marketing Department.

Legal Framework

Tanzania has a legal framework that guides the collection and dissemination of official statistics. The legal provisions include the Statistics Act of 2015, revised in 2019 (cited as the Statistics Act CAP 351 R.E 2019), and the data dissemination policy. The two documents are readily available to the public and can be viewed or downloaded from the NBS website (The United Republic of Tanzania, 2010 & 2019). The Statistical Act CAP 351 R.E 2019 stipulates the principles guiding the collection and dissemination of statistical information, which should abide by international standards. It mandates the NBS to coordinate and supervise the country's NSS, placing it in charge of the production, coordination, and dissemination of official statistics, according to the principles stipulated in the African Charter on Statistics as adopted from the UNFPNOS (African Union, 2000). As such, the NBS is the sole custodian of national official statistics. Furthermore, as part of this act, a controversial amendment was passed in 2018, placing criminal liability on anyone disseminating statistics that contradicted the official government statistics or failed to follow national and international statistical standards. However, the clause was removed in the 2019 amendment following contention by human rights activists and civil society (Fosci et al, 2019). Nevertheless, Article 33 of the Statistics Act CAP 351 R.E 2019 still protects the identity of respondents by anonymisation and confidentiality on the part of the data users to whom the statistical records are disclosed. This demonstrates the delicate balance countries must strike when striving to build data sharing systems whilst also protecting the respondent's confidentiality.

The pricing and dissemination policy sets out the principles that govern the dissemination of official statistics and microdata. Among other principles, it states that anonymized microdata will be available for scientific purposes and that data produced by NBS is free. However, a data user may be charged a fee if further processing of the microdata is required. Additionally, there is limited information on the country's actual practice and progress on microdata dissemination because the online repository is largely inaccessible due to the ongoing remodelling.

Microdata Accessibility

In Tanzania, some surveys are more likely to be disseminated than others; for example, the National Panel Survey, the Household Budget Survey, and the Demographic Health Surveys are commonly requested by users. This is due to their regularity; hence it is possible to measure trends in the population over time using these surveys. Additionally, these surveys are either fully or partially supported by international partners therefore this microdata can also be accessed externally on international organisations' repositories. Where sampling is deemed representative of the national population, surveys are disseminated. On the other hand, for surveys with small sample sizes or covering a small geographical area, data is used internally and not disseminated to the public; however, data users may still request access to this microdata through the Statistician-General. NBS states it has disseminated microdata for more than 75% of the surveys carried out in the last 10 years, although there is no way of independently verifying this information.

Microdata is not widely disseminated on the NBS website or the microdata portal (TNADA). Currently, microdata is only disseminated for the Household Budget Survey 2017-2018, National Panel Survey 2010-2011 & 2012-2013, and Agriculture Sample Census Survey 2002-2003, which are formatted as PUFs on the TNADA. Nevertheless, this link on the NBS website to the microdata archive is not working as NBS is currently restructuring the portal. In this repository, 30 surveys and their metadata in the format DDI/XML and JSON have been uploaded but microdata for the remainder 27 of these files are not yet available. NBS hopes to complete work on the repository by the end of March 2022 so that it is ready for public use. Notably, the NBS website includes a comment form embedded in the 'about us' section where data users can give feedback on accessibility to data and the general functioning of the website. This user feedback will be vital in improving the website and data availability in the dissemination process.

Once available, Tanzania's Dissemination and Pricing Policy notes that the repository will provide a straightforward procedure for microdata dissemination, following the UNSD microdata dissemination best practices by including tiered file types, namely PUFs, licensed files, and non-accessible data, with differing availability to users. In the meantime, microdata is disseminated upon request (CCSA, 2014). Data users can access microdata by contacting the Statistician General through the email address available on the website. From here, the Statistician General liaises with the IT department and the data release committee to release microdata to users who have met the conditions specified in the data request form (Dissemination and Pricing Policy, 2010).

The NBS upholds the principle of appropriate microdata use by requiring users to state their intended use of microdata, a requisite even for the PUF. Although microdata in Tanzania is mainly used for research, academic purposes and by the private sector, no academic qualification is required to access microdata. Instead, the NBS ensures equitability by making microdata available impartially and simultaneously to all data users, a stipulation of the African Charter of Statistics (African Union, 2000). However, access to microdata may still be denied if the reason given is for evaluating the ethnicity or religious affiliations of the population. Applications are authenticated by either validating the emails or calling the institutional contacts provided. This vetting process is not required if the request is from another government department.

Tanzania's microdata is further accessible through external repositories hosted by international organizations (see Appendix F). Whilst some external repositories, e.g., Food and Agriculture Organization of the United Nations (FAO), redirect you to the NBS website to access microdata, others have highly anonymized PUFs that are readily available after the user registers and logs in. As part of this, some repositories require an application form to be filled out when requesting access to licensed data sets.

Data Anonymity and Protection

Microdata is anonymized by removing direct and indirect identifiers. For example, local suppression is used to prevent the identification of respondents that are outliers in a relatively homogeneous group. To ensure confidentiality, as stipulated in Tanzania's 2017 statistics regulation, Act, the data users' output is reviewed by the Statistician-General or a mandated NBS officer before it is published. NBS does not allow the publication of data users' output that does not meet the requirement of data confidentiality. Regarding security, the NBS IT department works in conjunction with the government agency responsible for cyber security to ensure the repository's safety. To date, there has never been a breach of the NBS data security system, however the microdata dissemination portal (TNADA) has recently crashed due to Local Area Network issues of the host, the Tanzania e-Government Agency.

Vietnam

Data Dissemination Landscape

The General Statistical Office (GSO) of Vietnam was established through decree 23/CP on March 23rd, 1994. The original mandate for the GSO was to provide quantitative information on the socio-economic situation in Vietnam. However, the structure used to disseminate data before 2020 resulted in poor coordination between government agencies and ministries. With the increased requests for data by users such as ministries and agencies, researchers, students, and business professionals, the need for improvement in the statistical dissemination has been acknowledged by the Vietnamese government. In response to this, Vietnam has released a new strategy, the Vietnam Statistical Development Strategy 2021 - 2030, vision to 2045 (VSDS), outlining national goals seeking to improve the GSO's data dissemination capabilities. As such, the data dissemination landscape in Vietnam is currently being redeveloped.

The Department of Data Collection and IT Application, which is under the purview of the General Statistics Office, was formed as a response to the VSDS. In the department, there are 130 people with 55 people focusing on data collection and data management of 30 surveys each year. Currently, there is no unit that focuses on data dissemination; however, the department is hoping to establish one soon. Training of the staff regarding data dissemination is non-existent though staff are trained in other aspects of data collection. While changes within the department have been slow since its inception, the VSDS aims to address the lack of infrastructure, regulations, and procedures around the collection, management, and dissemination of collected data.

Legal Framework

Since the inception of Vietnam's GSO in 1994, the legal environment surrounding statistics and dissemination has evolved. Following the enactment of Law No. 89/2015/QH13, Vietnam's Law on Statistics, which became effective in 2016, there have been two subsequent legal documents produced specifically in relation to practices surrounding statistical data. Decision No.10/2020/QD-TTg which specifies the functions, duties, powers, and organizational structure of the GSO became valid in 2020. This legal document focuses on the coordination and sharing of statistical information between the GSO, ministries, and line ministries within Vietnam. The following year, Decision No. 2014/QĐ-TTg became valid. This decision addresses the approval of the VSDS. Furthermore, Vietnam aims to improve the production, provision, and dissemination of data through the Fourth Industrial Revolution, progressing the quality of human resources, science, and technology through innovation and digital transformation. Vietnam's goal is to achieve the advanced level in the ASEAN sub-region by 2030 and aim to have a statistical system as modern as other countries in the world by 2045 (VSDS, 2021).

The legal environment is still being developed. Although there have been no known cases of outputs not abiding by Vietnam's privacy laws, there is no system set up to review data users' output before it becomes public. In addition, there is no agreement that outlines responsibilities and repercussions of using and misusing data that must be signed by the user when they are provided with data. As of 2021, there is a renewed focus on creating a robust, synchronous, and timely legal environment, with the goal of amending the current Statistics Law as well as all related documents to ensure the independence of the proposed centralized statistical system (VSDS, 2021).

Microdata Accessibility

At present there is no formal process for requesting and disseminating microdata from the GSO. However, requests for data access can be sent by email, by phone, or through the GSO website. When requests are sent, a formal letter from the person or organization must be included which explains what the data will be used for. The request is then reviewed by the ministry holding the data. As the data is not centrally stored, if the request is accepted, data is disseminated in the format chosen by the ministry that is storing the data. As such, data may be sent in a variety of formats including using a USB or downloaded from the ministry website. Although there is a process for reviewing and disseminating microdata, there are no formal rules and regulations associated with the process of dissemination. As of 2021, the Department of Data Collection and IT Application has begun collecting data from the ministries and saving it to a physical server in their office with the goal of centrally storing all survey data. The department has not had the time to check to see whether the data they have received is good or not, therefore they are unsure if they are currently providing good quality data to the user. This is something they are looking to address. When accessing data there is a succession of users. Ministries have priority, then those who are in academia, followed by international organizations. In addition, the proposed use of data is ranked, with policymaking being the priority, followed by research and business.

The GSO website displays a collection of aggregated data, but they are variable specific. To access the microdata, a variable must be chosen in addition to the year and other specified aspects of the variable (i.e. gender, urban or rural, province). The data are then presented in a table format that does not support downloading the data or copying the table. Nevertheless, outside of the GSO, there are other organizations that allow the downloading of specific surveys (see Appendix F). However, these organizations only post surveys conducted within their related sector and there are a limited number of surveys available. It is important to note that the VSDS specifies a dissemination strategy that closely aligns with the Dissemination of Official Statistics chapter in the most current version of The Handbook on Management and Organization of National Statistical Systems (2021).

Data Anonymity and Protection

Before the dissemination of data, the GSO follows Statistical Disclosure Control techniques for anonymising microdata files. A common technique used by the GSO to anonymize data is non-perturbing-masking, where identifiable variables, such as name and address, are removed from the data (Dupriez & Boyko, 2010). For surveys requested by companies that includes enterprise data, tax codes are also removed. Currently, indirect identifiers such as birthdates and postal codes are not removed, and respondent characteristics are not re-coded. In addition, the practice of adding some level of "noise" to the data is not used. In the interview the need for an improvement in cleaning data before dissemination was acknowledged.

In terms of data protection, the currently the data is stored either on a server in the Department of Data Collection and IT Application or by the associated ministry. As mentioned previously, the department is still receiving microdata files from ministries and aim to have a complete collection of all surveys on the sever located in the Data Collection and IT Application Department. The server is and will only be used for storage, not dissemination purposes. Within the last few years, the GSO has invested in a security system (it has not been clarified whether this is a physical or technical system), however this is only valid for the next two years. Nevertheless, a request to reinforce and improve the present security system has been accepted by the government. Due to the goal of centralizing the data dissemination system and providing web-based access to data, the GSO recognizes that additional security measures will be required.

Discussion

Microdata Dissemination an Evolving Data Ecosystem

Microdata dissemination has become increasingly important as more countries are trying to disseminate data for policymaking and research purposes. The four countries in our report have employed practices to disseminate substantial amounts of data, with the goal of ensuring confidentiality and anonymity. As discussed previously, in 2014 the UNSD created a document outlining best practices for microdata dissemination (CCSA, 2014). Within the document, two collections of best practices were highlighted, key principles and organizational best practices as well as technical standards and best practice, which were influenced by the 2007 Organization for Economic Co-operation and Development (OECD) Principles and Guidelines for Access to Research Data from Public Funding (referred to as OECD Principles and Guidelines) (ibid). Along with these best practices, further guidelines and principles have been produced by multiple organizations, including Open Data Charter (ODC), Open Data Inventory (ODIN), UN General Assembly (UNGA), Open Data Barometer (ODB) and the International Household Survey Network (IHSN). For example, the ODIN measures a country's dissemination practices with respect to the international standards of openness as well as provides guidelines for countries to follow. As of 2020, Palestine had a score of 72 followed by Costa Rica at 58, whilst Tanzania and Vietnam had scores of 56 and 49 respectively (ODIN, 2020). These reports recommend practices to support lower risks of microdata dissemination and ensure technical, legal, and ethical issues are properly addressed (ibid). Through this research it has become clear that there are discrepancies between these recommended practices and the current technical and organizational practices of NSO's.

Although all four countries are at various stages, commonalities exist regarding the challenges the NSOs face, highlighted in the following four themes: funding, capacity to monitor use, microdata accessibility, and capacity and resources to train staff. While looking for guidance as to how countries are supposed to address these issues, many of the guidelines and reports provide vague explanations or reactive solutions.

Funding

Unfortunately, NSOs in many countries continue to remain underfunded (UNDESA, 2021). While the UN recommends that the annual budget allocation for NSOs be within the normal confines of the original government budget proposal or request, this is not always the case (2021). Many NSOs in LMICs rely on funding from donors and organizations to maintain essential parts of their operation. Albeit useful, this is not a long-term solution. Firstly, external financing in fringes on the control of NSOs with regards to their operations, compromising the design and implementation of a statistical program reflective of national priorities (UNDESA, 2021). Secondly, data infrastructure and digital operations require continual updates to maintain and enhance statistical capacity-building. In other words, sustaining a dissemination system that mirrors a best practice approach requires dedicated budgetary planning and continued funding - an ongoing challenge for three of these NSOs (CCSA, 2014). Vietnam and Tanzania are currently experiencing financial constraints as they try to advance their data dissemination practices, whilst Costa Rica has previously tried to address the issue of a decreasing budget.

As mentioned earlier, Vietnam is undergoing a major transition within the GSO. In addition to meeting the increased need for accessible data, they are consolidating and streamlining dissemination practices. However, budgetary restraint is hindering progress and the ability to revitalize data dissemination. The continuation of the daily running of the GSO, alongside the creation of a new data dissemination platform and the analyzing, cleaning, and anonymisation of 30 surveys yearly, demands increased funding to effectively perform and achieve best practice. Whilst these aims are positive, the process of achieving them is slowed because of financial constraints.

In Tanzania, according to the Statistics Act No.8 of 2018 s.21, the NBS budget should consist of: funds appropriated by Parliament; revenue from NBS products and services; sums that are borrowed, received, or made available to the NBS, as well as donations, grants, and bequests. However, financial constraints still exist, resulting in a lack of adequate microdata dissemination. For Costa Rica, Law No. 9694 ensures sustainable financing of the INEC's production and dissemination of statistics through funds from the annual government budget, along with an annual transfer from the Central Bank of Costa Rica and revenue from INEC products and services. Although Costa Rica has consistent funding and comprehensive dissemination practices in place, additional funds are still needed for: the development of new statistical strategies; statistical production and dissemination strengthening to ensure long-term sustainability, and the coordination of statistical activities within the NSS (OECD, 2019).

When looking through the available guidelines, mentions of funding and financial resources are limited. Whilst implementing technical best practices has a financial burden on NSOs, the CCSA (2014) argues that these financial challenges can be significantly reduced by adopting common tools and international guidelines. In the UNSD (2014) best practices, funding is only mentioned regarding entities outside of the government, or that maintaining advised practices requires budgetary planning and funding. The UNFPOS (2014) does not mention funding outside of Principle Five, which dis- cusses data sources. The OECD Principles and Guidelines also note the cost of data dissemination, specifying that dedicated budgeting and financial support are required but do not go into further depth as to how to achieve this (2007).

From the documents reviewed, the IHSN is the most informative (Dupriez & Boyko, 2010). A chapter is dedicated to the discussion of the costs and risks of disseminating data and how best to address this (ibid). While informative, it suggests accepting additional funding from outside sources to meet needs, however this may result in loss of control over data.

Capacity to Monitor Use

A common observation is that countries are experiencing difficulties monitoring the use of data by beneficiaries requesting access. This is an area they are all seeking to improve, although at the time of this report, Costa Rica is the only country to have experienced a confidentiality breach. Following this incident, Costa Rica amended its microdata dissemination policy, however there is still room for improvement. For example, any citizen can create an account through the INEC website using a valid email address with no further identification required. In addition, while users must provide a general description of purpose and data use, requests are not authenticated by the INEC and the user's output is not reviewed. In Vietnam, data is disseminated through individual platforms or using a USB when the requested data is not available through an online platform. Once provided, there is no follow-up required, nor any pre-approval before the output is published. Additionally, users are not required to sign a legal document committing to appropriate data use and the acknowledgement that they must abide by Vietnams' 2003 Statistics Law. As Vietnam continues to work towards an online platform holding all ministries data, they are likely to face similar issues as Costa Rica. Due to the limitations on data dissemination monitoring, these NSO's do not have the abilities to reassure the absence of re-identification.

In comparison, Tanzania and Palestine have a system that allows them to check the documents before they are published. PCBS has what is referred to as a 'monitoring statistical system' which is used to monitor, document, and follow-up on the assessment and evaluation of issues in the country (PCBS, 2017). The NBS in Tanzania is mandated by the Statistics Act CAP 351 R.E 2019 to control the quality and dissemination of information. Additionally, Tanzania's 2017 Statistics Act Regulations specify that any statistical information obtained by the requesting party must be authorized by the Bureau before publishing or communicating, ensuring that the data meets the laid principles (The Statistics Act Regulations, 2017).

According to Principle Six of the UNFPOS (2014), data is to be used strictly for statistical purposes. In addition, Principle Four highlights that statistical agencies are to comment on any misuse or un-founded claims regarding provided data (ibid). However, this is difficult when countries do not have the capacity to monitor use. While the IHSN does discuss the misuse of data, the recommendations provided include the signing of a user-agreement that will allow for legal actions to be taken if a breach occurs (Dupriez & Boy ko, 2010). IHSN also refers to the OECD Principles and Guidelines, aligning data use under the transparency, legal conformity, and formal responsibility principles (2007). While making sure legal action will occur if a breach happens or data is misused, it is a reactive solution. There are no guidelines or suggestions that support proactive responses to prevent the publishing of incorrect data or data misuse.

Microdata Accessibility

According to the UN (2021) Handbook on Management and Organization of National Statistical Systems, the most important dissemination tool is the data portal. A data portal is a web-based platform that supports interactive data and metadata dissemination. It can include multiple databases that model-specific data types, allowing for a succinct way to present data according to its type, i.e. microdata, macro data or geospatial data, making these datasets open, machine readable, free to use, readily available, and updated, in accordance with the ODB (2018) suggestions. While a platform of this nature is the goal for many countries, it may be an unrealistic or difficult achievement due to budgetary, legal, and personnel constraints. All countries in this report acknowledged the need for an accessible microdata dissemination platform, however not all countries have achieved it. Regarding dissemination of data, the IHSN guidelines suggest segregating data into PUFs, licensed files, and non- accessible data. However, it contrasts with the guidelines found in the ODC (2018) that emphasizes unrestricted licenses under Principle Three. Yet, the UNFPOS (2014) specifies that open data should not be equated with unrestricted data. In this regard, tiered data accessibility is hindered due to the lack of clarity and consensus concerning this aspect, complicating the comparison between NSOs prevailing practices with these inconsistent recommendations. Hence, questioning the simplicity of the ODIN openness scores mentioned previously.

Currently, Vietnam does not have an online repository. Before the VSDS, data was disseminated by the ministry or provincial statistical office that organized, implemented, or funded the survey. Although the GSO facilitates the dissemination of data, there is no streamlined approach resulting in a convoluted process for dissemination. In addition, practices do not follow any rules or official guidelines and the GSO relies on a letter of intent to determine whether they will release data to the requested party. To advance the accessibility of microdata dissemination, in 2021 the VSDS provided new actionable goals in alignment with the progressing consolidation and the creation of a data portal which is currently underway. With the anticipated release of this database being within 2022, Vietnam hopes to meet the needs of the increasing demand for accessible data. However, Vietnam is in the process of determining the best way of presenting data sets whilst protecting the data within the new portal and is looking for additional insight into this matter.

Tanzania is also undergoing remodelling of its online microdata repository. While the country has very few data requests, it hopes an accessible microdata repository will increase user's awareness of existing statistical data documentation. In Tanzania, access is made available in order of importance by the requesting party, from ministries to academia to international organizations. The NBS did reference plans of segregating data into PUFs, licensed files and non-accessible data, in line with the IHSN (2010) guidelines, to maintain anonymity of data, yet this cannot be confirmed until the remodelled repository is launched. Instead, there are various links on the NBS website that redirect users to specific databases. Each database provides the data in different formats with some including quick statistics for various surveys in the form of graphs, maps, tables, and charts. For the Tanzania Socio- Economic Database (TSED), an accompanying application has been created and is accessible on android and iOS devices. Although the application supports easy access to aggregated data, it does not allow for the dissemination of microdata. However, Tanzania may develop the application further to expand microdata accessibility in the future.

Members of PCBS stated that Palestine has disseminated approximately 95% of its data within the last 10 years. While it is easy to download surveys from the website, to access microdata a request must be sent through the website. Once accepted, a link is provided to allow the user to download the disseminated data, all of which is anonymized. Even though substantial amounts of Palestine's microdata are accessible, it aims to increase the availability and accessibility of statistics further as stated in sub-objective 1.2 of the 2018-2022 National Strategy for the Development of Official Statistics (NSDS).

Out of the four countries, Costa Rica has the most user-friendly, readily available, and accessible database. As a point of comparison, Costa Rica has followed the UNSD guidelines by providing an accessible online repository with tiered microdata disclosure respecting confidentiality. However, the ODB (2018) indicates that Costa Rica has not made much improvement with respect to open data accessibility since it signed the G-20 Anti-Corruption Open Data Principles. Thus, despite complying with best practices, all countries have further scope for improvement.

Capacity and Resources to Train Staff

With an increase in the sophistication of data use, keeping staff up to date with improved anonymization techniques, dissemination guidelines, and the evolving legal environment through capacity training is important. However, such training is expensive and with limited budgets, training is repeatedly neglected. Unfortunately, some countries do not provide PUFs as they lack the capacity for thorough anonymization (Winkler, 1998; UNDESA, 2021). The challenge of capacity and resource training for staff is recognised by all countries in this report as an area that needs improvement.

Firstly, Tanzania discussed that the NBS faces challenges in making microdata available to the public due to knowledge gaps among staff. Here training surrounding how to package data and how to remove identifiers has been identified as an area that is lacking and where training would be beneficial. Nevertheless, resources are available, and staff are encouraged to participate in capacity building training that is held internally or in collaboration with other organizations. For example, staff at the NBS can apply online to attend an eight-day STATA course tailored to their level of understanding, however this only occurs if the budget allows.

As Vietnam is currently undergoing infrastructure restructuring and consolidation of its dissemination practices, its focus is on the creation of a usable and accessible data portal as well as the provision of good data. Vietnam does not have any formal training for staff regarding microdata dissemination practices. The staff are currently provided training in the collection of data and other data related practices. Although Vietnam acknowledged that the goal is to provide further training focusing on data dissemination best practices, it is currently not a top priority. In Palestine, the PCBS does not provide training to their staff; even though, in the NSDS there is mention of using training programs to support the processing of statistical data from administrative records. Additionally, the NSDS highlights the provision of training courses for statistical staff within the NSS as an important activity to support the sub-objective of an improved work environment. However, there was no clear answer as to when the training will be implemented.

Similarly, Costa Rica does not currently provide formal training on microdata dissemination. Based on amended 2019 law, there is a general policy and online guidelines on how to disseminate data, yet there is no policy to support the training of INEC staff. Upon hiring, individuals are informed about these guidelines and their duty regarding the confidentiality of personal information; however, based on Law No. 7839, INEC staff are not required to sign formal confidentiality agreements and there is no training to accompany this information (OECD, 2019).

In the list of organizational best practices, points regarding the quality of data, data security, and the legal environment surrounding data dissemination are unlikely to be properly met when training of staff does not occur (CCSA, 2014). It is specified in the UNSD best practices that the onus is on data managers and data collection organizations to make sure explicit quality standards are followed (ibid). However, without continuous and reliable training for staff there is no confirmation that the staff are consistent in their practice. For security, the focus is on the use of techniques and instruments to guarantee the integrity of data; therefore, without training on current best practices, it is difficult to guarantee secure data. A similar scenario is evident for legal frameworks and the best practices surrounding access to, and the dissemination of, data. Additionally, data anonymization techniques have become increasingly important and difficult due to the availability of geospatial information and data integration (Croft et al., 2021). Without training to ensure proper compliance with evolving practices regarding data anonymization, the data may be identifiable. While IHSN mentions that producing good metadata can assist in training new staff, training new users in the use of microdata files is the focus (Dupriez & Boyko, 2010). Otherwise, training of NSO staff is not mentioned in any of the reviewed guidelines or reports.

Conclusion

Wide sharing of microdata is the most effective way of leveraging microdata's analytical potential and maximizing the return on investment in household surveys. This report explores the state of microdata dissemination for household surveys across four countries to compare the prevailing practices with the recommended practices. The report further explores emerging challenges, opportunities for future development, and areas where supporting NSOs with microdata dissemination is needed. Generally, the countries have legislative policies that guide the collection and dissemination of national statistics and microdata, although the degree of execution varies. Nevertheless, these laws and policies have been amended over the years to reflect the UNFPOS (2014). These amendments demonstrate countries' commitment to improving the quality of official statistics and the standardization of their microdata dissemination procedures. All countries are disseminating microdata but at various levels. Whilst Costa Rica and Palestine have information available on their online microdata repositories, Vietnam and Tanzania are currently in the process of developing a functional and accessible online repository for widespread microdata dissemination. Despite this variation, all countries are conscious of data anonymization and protection, with no breaches in data security systems and confidentiality agreements being reported since the enforcement of effective legislation.

Respectable microdata dissemination involves the accessible provision of data and its related documentation with conditions attached outlining how the data can be used (Dupriez et al, 2010). Yet, despite the availability of manuals, handbooks, and policies with fundamental principles, technical standards, and best practices for disseminating microdata, the four countries still face various challenges in effectively implementing them. This report identified the following challenges:

1. **Funding:** Underfunding of NSOs by the national budget and overreliance on inconsistent and unreliable external funding.
2. **Monitoring Use:** Limited scope in monitoring microdata uses once it is disseminated.
3. **Accessibility:** Some countries lack online microdata repositories while others have repositories that are not fully functional. Access to microdata for such countries is constrained to making requests to the NSOs via email that may be time-consuming; and
4. **Training:** Limited capacity and resources to train staff on microdata anonymisation and dissemination.

Recommendations

To address the challenges outlined above, we propose recommendations to increase collaboration, technical support, funding, and user feedback as well as measures to ensure confidentiality through the standardization of procedures reviewing data user's output.

Capacity Building

Training

Collaboration is required between development partners and countries to enhance training and benchmarking for NSOs, as well as the sharing of knowledge on microdata anonymisation, packaging, dissemination and any additional technical techniques.

Consistent Guidelines

Interagency coordination to produce an abridged version of microdata dissemination guidelines that are universally consistent in regard to the security, access, legal frameworks, and the technical aspect of dissemination to provide a streamlined and cohesive approach for countries.

Resources

Funding

Countries require low-cost options for data portals. In addition, highlight the cost-effectiveness of abiding by the best practices guidelines in a tangible format to encourage compliance and to encourace governments and other development partners to invest in NSOs.

Data Security

Standardization of dissemination guidelines and best practice recommendations regarding how to monitor data users' output to promote confidentiality and appropriate use of provided data.

Technical Support

The development or increasing the ability to access user-friendly data portals that are embedded with microdata request forms and metadata download functions to guarantee ease of access, as is the case with international organization's repositories

References

African Union. 2000. African Charter on Statistics. Available: https://www.nbs.go.tz/nbs/takwimu/ACT/ AFRICAN_CHARTER_ON_STATISTICS.pdf

Archivo Nacional Costa Rica. 2022. Contact Us. Available: https:// www.archivonacional.go.cr/index.php/ contactenos Accessed: 24th February 2022.

Croft, T., Dontamsetti, T., Mayala, B., Blankespoor, B., & Murray, S. (2021). Spatial Anonymization (p. 26). Committee for the Coordination of Statistical Activities. 2014. Microdata dissemination best practices. UNSD.
Available: https://unstats.un.org/unsd/accsub/2014docs-23rd/SA-2014-4-Microdata.pdf

Duncan. G. T., Sallie. K. A,, Stokes. L. S. 2001. "Disclosure Risk vs. Data Utility: The R-U Confidentiality Map.", 2001Technical Report, U.S. National Institute of Statistical Sciences.

Dupriez. O., Ernie. B. 2010. "Dissemination of Microdata Files. Formulating Policies and Procedures", International Household Survey Network, IHSN Working Paper No 005.

Elkatawneh. H. 2016. Five Qualitative Approaches: Problem, Purpose, and Questions/The Role of Theory in the Five Qualitative Approaches/Comparative Case Study. SSRN.

Gomatam. S., Karr. A. F., Reiter. J. P., Sanil. A. P. 2005. Data Dissemination and Disclosure Limitation in a World without Microdata: A Risk-Utility Framework for Remote Access Analysis Servers. Statistical Sci- ence, 20(2), 163–177.

Hyett. N. Kenny. A., Dickson-Swift. V. 2014. Methodology or method? A critical review of qualitative case study reports. International journal of qualitative studies on health and well-being, 9, 23606. https:// doi.org/10.3402/ qhw.v9.23606

Information Commissioner's Office. 2012. Anonymisation: managing data risk protection code of practice.
Available: https://ico.org.uk/media/1061/anonymisation-code.pdf

Innovative Routines International. 2022. Anonymizing Indirect Identifiers to Lower Re- ID Risk. Available: https://www.iri.com/blog/data-protection/anonymizing-indirect-identifiers-lower-re-id-risk/

Johnson. W. P. 2014. Snowball Sampling: Introduction. Wiley StatsRef: Statistics Reference Online, 2014.

National Institute of Statistics and Censuses. 2014. Disclosure Policy of Statistics and Delivery of Microdata of the Institute National Statistics and Census.

Noble. H., Smith. J. 2015. Issues of validity and reliability in qualitative research. Evidence-Based Nursing 2015; 18:34-35.

National Information Standards Organisation. 2004. Understanding metadata. Available: http://www.niso.org/ publications/press/UnderstandingMetadata.pdf

Organisation for Economic Cooperation and Development (OECD). 2007. "OECD Principles and Guidelines for Access to Research Data from Public Funding". www.oecd.org/dataoecd/9/61/38500813.pdf

Organisation for Economic Cooperation and Development (OECD). 2019. OECD Assessment of the Statistical System and Key Statistics of Costa Rica. Available: https://www.oecd.org/statistics/good-practice- toolkit/ OECD-Review-of-the-Statistical-System-and-Key-Official-Statistics-of-Costa Rica.pdf

Palestinian Central Bureau of Statistics. 2006. Code of Practice. State of Palestine. https://www.pcbs.gov.ps/site/lang en/575/default.aspx.

Palestinian Central Bureau of Statistics. 2006. National Strategy for Development of Official Statistics 2009-20.
State of Palestine. Available:
https://www.pcbs.gov.ps/Portals/_Rainbow/Documents/NSDS_e.pdf
Palestinian Central Bureau of Statistics. 2017. National Strategy for the Development of Official Statistics 2018-2022. State of Palestine.
Palestinian Central Bureau of Statistics. 2018. Selected Indicators for Internal Trade Activities in Palestine.
Available: https://www.pcbs.gov.ps/Portals/_Rainbow/Documents/Trade_annual2_2018e.html
Palestinian Central Bureau of Statistics. 2019. Dissemination Policy Long Version. State of Palestine. Available: https://www.pcbs.gov.ps/Portals/_Rainbow/Documents/PCBS-SHORT%20VERSION-Dissemination% 20Policy.pdf
Palestinian Central Bureau of Statistics. 2021. Dissemination and Documentation DepartmentPrivacy Policy.
State of Palestine. Available: https://www.pcbs.gov.ps/Document/pdf/privacy-policy-english.pdf
Palestinian Central Bureau of Statistics. 2022. Legal Basis. State of Palestine. https://www.pcbs.gov.ps/site/ lang en/539/default.aspx
Palestinian Central Bureau of Statistics. 2022. Pcbs.gov.ps. PCBS [online]. Available: https://www.pcbs.gov.ps/site/lang en/1/default.aspx [Accessed 27 February 2022].
Palinkas. L. A., Horwitz. S. M., Green. C. A., Wisdom. J. P., Duan. N., Hoag Wood. K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. Administration and policy in mental health, 42(5), 533–544. https://doi.org/10.1007/s10488-013-0528 -y
Patton. M. Q. 2015. Qualitative Research & Evaluation Methods: Integrating Theory and Practice.
Reiter. J. P. 2012. Statistical Approaches To Protecting Confidentiality For Microdata And Their Effects On The Quality Of Statistical Inferences, Public Opinion Quarterly, Volume 76, Issue 1, Spring 2012, Pages 163– 181, https:// doi.org/10.1093/poq/nfr058
Robinson, O. C. 2014. 'Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide', Qualitative Research in Psychology, 11(1), pp. 25–41. doi:10.1080/14780887.2013.801543.
Ruggles. S. 2014. Big Microdata for Population Research. Demography 1 February 2014; 51 (1): 287–297. doi: https://doi.org/10.1007/s13524-013-0240-2.
Stake. R. E. 1995. The art of case study research. Thousand Oaks, CA: Sage; 1995. Stat Strategy_2021- 2030_ENG.docx. (n.d.). Retrieved March 3, 2022, from https://lsecloud-my.sharepoint.com/:w:/r/ personal/e_l_campbell harry_lse_ac_uk/_layouts/15/Doc.aspx?sourcedoc=%7BE0F332FC-D5BD 423A- 83C4-DF6E70EDD73B%7D&file=Stat%20Strategy_2021- 2030_ENG.docx&action=default&mobileredirect=true
The United Republic of Tanzania. 2010. National Bureau of Statistics. Dissemination and Pricing Policy. Availa- ble: https://www.nbs.go.tz/index.php/en/about-us/policies-and-legislations/policies/12-dissemination- and-pricing-policy
The United Republic of Tanzania. 2017. The Statistics Act Regulations, 8 § 38 (2017). Available: https:// www.nbs.go.tz/nbs/takwimu/ACT/THE_STATISTICS_REGULATIONS_2017.pdf

The United Republic of Tanzania. 2019. The Statistics Act CAP. 351 R.E. 2019. Available: https://www.nbs.go.tz/index.php/en/act/522-the-statistics-act-cap-351-r-e-2019
United Nations. 2021. Chapter 10—Dissemination of Official Statistics. In The Handbook on Management and Organization of National Statistical Systems (4th Edition Handbook Statistical Organization). Available: https://unstats.un.org/wiki/pages/viewpage.action?pageId=62587030
United Nations. 2014. Fundamental Principles of Official Statistics. Available: https://unstats.un.org/unsd/ dnss/gp/FP-New-E.pdf
United Nations Department of Economics & Social Affairs (UNDESA). (2021). The Handbook on Management and Organization of National Statistical Systems (4th ed.). United Nations.
United Nations Economic Commission for Europe (UNECE) and Statistics Sweden. 2003. "Statistical Confiden- tiality and Access to Microdata. Proceedings of the Seminar Session of the 2003 Conference of European Statisticians". Available: http://www.unece.org/stats/publications/ statistical.confidentiality.pdf
United Nations Economic Commission for Europe (UNECE), Conference of European Statisticians. 2007. "Managing Statistical Confidentiality and Microdata Access: Principles and Guidelines of Good Prac- tice". Available: http://www.unece.org/stats/publications/Managing statistical confidentiality.and.microdata.access.pdf
United Nations Economic Commission for Europe (UNECE). 2009. "Principles and Guidelines on Confidentiality Aspects of Data Integration Undertaken for Statistical or Related Research Purposes". Available: http://www.unece.org/stats/ publications/Confidentiality_aspects_data_ integration.pdf
United Nations Statistics Division. 2014. Country Profile of Occupied Palestine Territory. Available: https:// unstats.un.org/unsd/dnss/docViewer.aspx?docID=686#start
Vietnam Statistical Development Strategy 2021—2030, vision to 2045, no. 2014/QĐ-TTg (2021). Willenborg. L. C. R. J., De Waal. T. 2001. Elements of Statistical Disclosure Control. Springer, New York.
Winkler., W. E. 1998. Re-identification methods for evaluating the confidentiality of analytically valid microda- ta. Research in Official Statistics, 50–69.
Wolff. E. W. 2015. Wealth Distribution. International Encyclopedia of the Social & Behavioral Sciences (Second Edition), 2015. Retrieved from https://www.sciencedirect.com/topics/social-sciences/household-survey

Appendix

Appendix A: Terms of Reference

Taskforce for household survey microdata dissemination May 2020

1.Rationale

Data collected through censuses, surveys, administrative systems, and other sources form the foundation of official statistics, and are an invaluable input to research and decision making. These data are commonly referred to as microdata, defined as unit-level information on people or entities (such as individuals, households, business enterprises, farms, or even geographic areas). They are aggregated and analyzed to generate national estimates by official statisticians and analyzed by researchers and policy analysts to gain scientific insights which can be translated into policy.

The power of microdata stems from its granularity. Because microdata contain unit-level information, they facilitate the investigation of the unique ways a certain phenomenon may affect different groups or sub- populations. Microdata can be costly to collect, and microdata files often contain many variables, more than any single organization can fully exploit. An effective way to maximize the return on investment and fully leverage microdata's analytical potential is to share them widely. The role of governments in promoting the wider use of microdata has become increasingly important, and many initiatives have been implemented with not only the goal of maximizing potential uses, but also in the interest of transparency and openness.

Over the past two decades, a consensus has formed that data, particularly data produced by public bodies, should be openly and freely available[1]. This is in keeping with principle 1 of the Fundamental Principles of Official Statistics (FPOS), which says, in part, "[O]fficial statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information." Several indexes have been created to measure the openness of government datasets[2] However, these datasets largely consist of aggregate indicators. Microdata, which often contain identifiable information of individuals or other entities, are governed by principle 6, which says, "Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes."[3] This places limitations on what the data can be used for and who can use it. The dissemination of microdata requires careful attention to the disclosure of confidential information.

The World Bank, the International Household Survey Network (IHSN), UNECE, and other organizations have published best practice guidelines, handbooks, and software tools to assist organizations in disseminating microdata. Additionally, a previous report on Standards and best practices for survey data documentation was produced for the CCSA4 in 2014. These documents identify the key principles, technical standards, and best practices for disseminating microdata. They have paved the way for international organizations and national statistical offices to develop microdata dissemination policies.

[1] For example, the International Open Data Charter (https://opendatacharter.net/principles/)

[2] For example, the Open Data Inventory (http://odin.opendatawatch.com/) and the Open Data Barom- eter (https://opendatabarometer.org/).

[3] https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx

[4] https://unstats.un.org/unsd/accsub-public/microdata.pdf

Despite the availability of these manuals and tools, many countries still do not disseminate microdata, and those that do don't always follow the recommendations. Furthermore, in recent years, the data ecosystem has changed dramatically. To improve timeliness and fill gaps between survey rounds, National statistical offices are increasingly looking to take advantage of new sources of data, such as remote sensing, devices that make up Internet of Things (IoT), and social networks. These new data sources are seen as possible cost- effective complements to sample surveys. New or higher frequency data may fill gaps which surveys are unable to fill quickly enough while at the same time surveys may provide ground truthing (and quality checks) for these new sources of data. From a privacy and quality perspective many of these new data sources provide additional challenges for microdata dissemination. For example, geo referenced data retained in microdata are needed to link with remote sensed data, which may breach respondents' confidentiality. Additionally, as more and more new data sources become publicly available the ability to link data and identify individuals may increase substantially. It follows that new practices of integrating new data sources with microdata will require some changes in data management and dissemination practices.

In this context, the task force aims to provide an overview of the state of the art in terms of microdata dissemination in countries comparing the prevailing practices with the recommended practices. Such a review will be carried out in a number of countries that are selected to represent those in different levels of statisti- cal development and in different geographical regions. Furthermore, due the changing data ecosystem, emerging challenges will be identified in order to set an agenda for future research and international efforts to promote microdata dissemination.

2.Proposed output(s)
The task force will result in a technical report consisting of four chapters. The report will follow the structure below:

Chapter 1: Introduction: Microdata dissemination in the era of Open Data – This chapter will provide the reader the context and motivation of the report. An introduction on the use and value of microdata, a brief history of international efforts to promote microdata dissemination, and a description of open data principles will motivate Chapter 2. A description of the changing data ecosystem and potential impacts on microdata dissemination and use will motivate Chapter 3.

Chapter 2: State of the art of microdata dissemination – The objective of this chapter is to provide an overall picture on the availability of microdata, and compare the prevailing practices in a set of countries varying in geography and statistical capacity with recommended practices. The group will leverage information already obtained by the World Bank, Open Data Watch, and FAO to estimate the availability of microdata across countries and statistical domains. To provide examples of current practice, a group of countries will be selected, as much as possible, representing those in different levels of statistical development and in different geographical regions. Comparisons will be made of their current practices and best practices defined in international manuals across the following aspects:

1. Enabling legal framework, licensing, and terms of use
2. Quality of metadata, assessed by their accuracy, completeness, and adherence to recognized standards such as the Data Documentation Initiative (DDI)
3. An assessment of statistical disclosure control (SDC) policies developed by NSO's (including methods applied, if publicly disclosed) as well as well the presence of well composed legislation around SDC of microdata.
4. Availability and accessibility of microdata (How countries choose which datasets to dissemimate, as well as IT, legal, and other constraints)

Whenever applicable to any of the items listed above, good practices of international agencies in disseminating survey microdata will also be discussed.

Chapter 3: Microdata dissemination in an evolving data ecosystem – The objective of this chapter is to identify new challenges and gaps in the existing manuals on microdata dissemination, and point to the way forward for the international statistical community to focus on. These can be organized based on the aspects analyzed in Chapter 2.
For example:

- Enabling legal frameworks: An analysis based on a review of a few key data privacy legislations. For example, using the GDPR as an example of how privacy legislation impacts microdata dissemination in Europe.
- Quality of metadata: Enabling data interoperability through detailed standards-based metadata.
- Statistical disclosure control: Additional SDC considerations presented by alternative data sources
- Availability and accessibility: Exploring the feasibility of trusted repositories, remote access and virtual enclaves in low- and middle-income countries as a means of overcoming barriers to microdata dis- semination. This includes an assessment of sustainable funding, technical requirements, legal requirements.

Chapter 4: Conclusion (summary) – The objective of this chapter is to summarize (a) current status of microdata dissemination, highlighting issues and challenges faced; (b) recommendations on various approach- es and tools that work for countries at different capacities; and (c) areas for future work.

Appendix B: Information Sheet

Taskforce For Household Survey Microdata Dissemination

Researchers: Christine Imali, Mahima Chaudhari, Lindsay Caird and Emily Campbell-Harry
MSc Students in the International Development Department at the LSE

Information for participants

Thank you for considering participating in this study which will take place from 1st November 2021 – 25th March 2022. This information sheet outlines the purpose of the study and provides a description of your involvement and rights as a participant if you agree to take part.

1. What is the research about?

This research will be conducted as part of an external postgraduate consultancy project for the Inter-Secretariat Working Group on Household Surveys (ISWGHS) in the United Nations Statistics Division (UNSD). The goal of the research is to better understand the current practices, challenges countries face when disseminating household survey microdata. The research findings will be produced by us (the LSE postgraduate consultancy team) and published by the UNSD as part of a wider research project with the goal of supporting country offices in disseminating microdata more efficiently and securely.

2. Do I have to take part?

It is up to you to decide whether to take part. You do not have to take part if you do not want to. If you do decide to take part, we will ask you to sign a consent form that you can sign and return in advance of the interview or at the interview.

3. What will my involvement be?

You will be asked to take part in an interview about your experience and knowledge of the process by which household survey microdata is disseminated in your country. Additionally, the interview/ questionnaire may cover why or why not household survey microdata is disseminated, the security measures involved in sharing the data and other questions that align with the research topic. The interviews should take approximately 60 to 90 minutes.

4. How do I withdraw from the study?

You can withdraw from the study at any point until 5th February 2022, without having to give a reason. If any questions during the interview or in the questionnaire make you feel uncomfortable, you do not have to answer them. Withdrawing from the study will have no effect on you. If you withdraw from the study, I will not retain the information you have given thus far, unless you are happy for me to do so.

5. What will my information be used for?

We will use the collected information for research publication focusing on the dissemination of household survey data by country. Your information will be used by the LSE research students named above for research purposes and the creation of the report. The data may be used by the United Nations Statistical Division.

6. Will my taking part and my data be kept confidential? Will it be anonymised?

The records from this study will be kept as confidential as possible. Your name and personal information will not be used in any reports or publications resulting from the study without your consent. The name of the country you are representing will be mentioned in the final report. The digital files and transcripts of the interviews will be retained by the United Nations Statistics Division for the duration of the study and protected in accordance with United Nations policies on confidential information.

Limits to confidentiality: Confidentiality will be maintained as far as it is possible unless you tell us something which implies that you or someone you mention might be in significant danger of harm and unable to act for themselves; in this case, we may have to inform the relevant agencies of this, but we would discuss this with you first.

7. Who has reviewed this study?
This study has undergone an ethics review in accordance with the LSE Research Ethics Policy and Procedure.

8. Data Protection Privacy Notice
The LSE Research Privacy Policy can be found at:
https://info.lse.ac.uk/staff/divisions/Secretarys-Division/Assets/Documents/Information-Records-Management/Privacy-Notice-for-Research-v1.2.pdf?from_serp=1

The legal basis used to process your personal data will be a Public task. The legal basis used to process special category personal data (e.g. data that reveals the racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health, sex life or sexual orientation, genetic or biometric data) will be for scientific and historical research or statistical purposes.

To request a copy of the data held about you please contact glpd.info.rights@lse.ac.uk

9. What if I have a question or complaint?
If you have any questions regarding this study please contact the researcher, Mahima Chaudhari, on M.Chaudhari1@lse.ac.uk.

If you have any concerns or complaints regarding the conduct of this research, please contact the LSE Research Governance Manager via research.ethics@lse.ac.uk.

If you are happy to take part in this study, please sign the consent sheet attached/below.

Appendix C: Consent Form

Taskforce For Household Survey Microdata Dissemination

Researchers: Christine Imali, Mahima Chaudhari, Lindsay Caird and Emily Campbell-Harry

PARTICIPATION IN THIS RESEARCH STUDY IS VOLUNTARY

I have read and understood the study information dated [DD/MM/YY], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	YES / NO
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and that I can withdraw from the study at any time up until February 5, 2022, without having to give a reason.	YES / NO
I agree to the interview being audio/video recorded.	YES / NO
I understand that the information I provide will be used for a research publication focusing on the dissemination of household survey data by countries and will be used by the LSE research students named above. The data may also be used by the United Nations Statistical Division.	YES / NO
I agree that my information can be quoted in research outputs.	YES / NO
I understand that any personal information that can identify me – such as my name, address, will be kept confidential and not shared with anyone other than the research team without my consent.	YES / NO
I give permission for the information I provide to be deposited in a data archive so that it may be used for future research.	YES / NO

Please retain a copy of this consent form.

Participant name:

Signature: _____ Date _____

Interviewer name:

Signature: _____ Date _____

For additional information, please contact Mahima Chaudhari at M.Chaudhari1@lse.ac.uk.

Appendix D: Interview Questions

Microdata Dissemination Interview Questions
Interview questions have been organized under the Five Safes, which is a framework used to make decisions about how to use data that is confidential or sensitive. Each of the Safes covered below provides a brief outline of the information we are trying to extract.

Date of interview:
Respondent's name:
Respondent's department:
Respondent's position:
Years in current position:
Respondent's email address (for possible follow-up):

Safe Project – This section is about the considerations undertaken by NSO when releasing the data to the user. Does it inquire as to what specifically the data will be used for? Are legal, moral, and ethical considerations taken into account when reviewing an application for data release?

- Do you require a formal letter from a research institution or an organization the researcher is representing as a requirement to access the data?
- How do you authenticate the validity of a user's request?
- Do you require information about the intended use of the microdata? If so, what are the primary reasons for requesting access?
- Have you ever denied a request to access data due to the reason provided for its usage? If yes, please provide an example?

Safe People – This section assesses the degree to which the NSO is aware of the type of people that microdata is made available to. Are there any identifiers in potential users' background information that makes them a security risk? Does the NSO collect new data types based on the changing patterns of users' needs? Is an assessment done to know what kind of data people require, which reinforces the NSO's decision on the data it needs to be collecting?

- Who are the potential users?
- Is the microdata available to the general public or only to qualified individuals?
- What are the qualifications required to access microdata?
- Do users go through a tutorial explaining how to use and store the data?
- Do users sign an affidavit or any legal documents before accessing data?
- Are there any legal ramifications for users who breach the data use agreement? If yes, what is it?

Safe Data – This section considers the safety and confidentiality of the data provider. What are the potentials for their identification within the data? Is a data user able to identify the provider through information in the data?

- Are all microdata files available for dissemination?
- How are the microdata documented (metadata)?
- How are the microdata anonymized?
  a. Which direct identifiers are removed (e.g., name and address)?
  b. Are direct, but unrecognizable identifiers (e.g., social security number) removed?
  c. Are indirect identifiers removed (e.g., birthdates, postal code)?

- Are respondent characteristics recoded to prevent identification (e.g., age, income, education level)? What types of recoding are employed?
- What other steps are taken to anonymize data and prevent reidentification?
- Are certain types of surveys more likely to be disseminated than others?
  Are there files that are considered for public use? Are there some files that are licensed and available for restricted usage?

Safe Setting – This section is about how data is accessed. What are the practical controls and channels available for accessing data?

- Is there a national archive or repository of government microdata?
  a. What is its name and address?
  b. Can users access the microdata in the national repository remotely?
  c. If no national repository, how are microdata disseminated?
- Are all microdata disseminated through the national repository or are some only available from international organizations?
- What is the policy for deciding whether to disseminate microdata domestically or through international organizations?
- What proportion of all surveys conducted in the last 10 years have been disseminated as microdata online?
- If survey microdata is not disseminated online, what are the reasons (technical, financial, legal, or others?)
- Do you have a data security system?
- Have you had any breaches in your security system? If yes, how was the breach managed?

Safe Outputs – This section is about safety systems in place to ensure that the data user follows the guidelines, and the output does not jeopardize data providers' privacy.

- Is there a system set up to review data user's output before it becomes public?
- Have there been instances where the output did not conform to privacy laws?
  a. If yes, what was the outcome, and how was the matter handled?
- Has there been any action taken by a data provider against the NSO or a data user because of information that became publicly available and revealed their identity?

Appendix E: Desk Research Outline

Microdata dissemination - Desk research

A list of NSOs with online microdata repositories ((Costa Rica appears to be the only country with an online presence that is functioning. Vietnam looks like it wants to start but is yet to populate. Tanzania says it does, but none of the links are working. Nothing for Mongolia. Palestine does not have an online presence, but I have attached the copy of a data release form I was asked to fill when I made an inquiry), out of the selected 5 will be made available. For those, with an online repository that is government or third-party sponsored:

- Where is it located? (Website name, URL)
- Who maintains it? (Name of government ministry/agency/department or third-party)
- What are its contents?
- Number and type of surveys
- Time period covered (date of surveys and date of posting to repository)
- Does the repository provide access to survey metadata (e.g. period data was collected, survey or administrative data, PAPI or CAPI etc)? If so, what is provided? In what format (e.g. Data Documentation Initiative (DDI); Dublin Core Metadata Specification (DCMI))?
- Are there microdata from other sources (census or administrative records) also stored in the repository? If so, what are they and where do they come from?
- Does the repository provide access to the microdata? If so, under what conditions? (Note different access conditions may apply to different surveys.)
- If it is possible to download microdata, what formats (ASCII, SQL) are provided?
- Is the microdata saved locally or is it available through a third-party website?
- Is there any information on how the surveys have been used?
- Is there information on the number of times the metadata for the survey have been viewed?
- Is there information on the number of times the microdata files have been downloaded?
- Look on the website of the survey sponsor for additional information on the purpose of the survey and its use

General observations on the ease of use of the repository:

- Was the repository and its user interface easy to use? How many steps did it take to get to the microdata from the 'Homepage'?
- Did the download function work?
- Was the website reliably available?

There may be more than one repository – if so, collect the same information for each one.

If the country does not have an online repository, it will be necessary to look for evidence of surveys that have been conducted over the last 10 years. This may be done by looking for published indicators and then seeking metadata that describe their sources. There may be information about survey programs posted on the nation- al statistics office website or on ministry (usually planning or economic) websites. We can also provide a list of national surveys mentioned in international databases (such as IHSN/ILO/FAO/UNICEF MICS/Global Health Data Xchange); look for information about these surveys on official national websites. This search can also include countries with repositories.

- What household surveys have been carried out in the last ten years by official agencies of the national government?

Record name of survey and purpose and year or years conducted.

Who (ministry, agency, department) sponsored the survey?

Are metadata describing the survey available?

Are the microdata available? If so in what format?

Is there evidence of how the surveys have been used?

Country	Int'l Org	Access to survey metadata?	Access to survey microdata?	Possible to download microdata?	If so, under what conditions?	Link
Costa Rica	Food and Agricultural Organization of the United Nations (FAO)	Yes	Yes	Yes	Format: Unknown For FAO produced: Researchers must fill out an application for access to licensed dataset and terms and conditions must be agreed, including: cannot identify individual entities, sold etc. For INEC produced: no microdata available.	https://microdata.fao.org/index.php/catalog#_r=&collection=&country=54&dtype=&from=1985&page=1&ps=&sid=&sk=&sort_by=popularity&sort_order=desc&to=2021&topic=&view=s&vk=
	Global Health Data Exchange	No	Yes	Yes	Where microdata is available, links to third party websites/external repositories are used (INEC; World Bank; IHME). These do not always work (eg IHME) and often a login is required to access the datasets on these websites.	http://ghdx.healthdata.org/geography/costa-rica
	International Labour Organization (ILOSTAT)	Yes	Yes	Yes	Documentation is provided on almost all surveys, with documentation being accessible through various PDFs. Surveys conducted in 2019 do not have data, even though on the results page for Costa Rica there is an icon indicating that data is available. However, there is locally stored microdata and the link for microdata access on INEC website but does not work.	https://www.ilo.org/surveyLib/index.php/catalog?sort_by=rank&sort_order=desc&sk=Costa+Rica#_r=1640868366184&collection=&country=54&dtype=&from=1975&page=1&ps=&sid=&sk=&sort_by=rank&sort_order=desc&to=2020&topic=&view=s&vk=
	International Household Survey (IHSN)	Yes	Yes	Yes	Where microdata is available, links to third party websites/external repository's are used (INEC; IPUMS International; World Bank; IHME). These do not always work (eg IHME).	http://catalog.ihsn.org/catalog#_r=&collection=&country=54&dtype=&from=1890&page=7&ps=&sid=&sk=&sort_by=nation&sort_order=desc&to=2021&topic=&view=s

	UNHCR	Yes	No	Yes	Format: SQI Must request access to available datasets, giving description of your research and agreeing to terms and conditions (including no attempt to identify respondents). Can only view or download once the application is accepted.	https://microdata.unhcr.org/index.php/catalog#_r=&collection=&country=54&dtype=&from=1991&page=1&ps=&sid=&sk=&sort_by=nation&sort_order=&to=2021&topic=&view=s&vk=
	UNODC	No	No	No		https://microdata.unhcr.org/index.php/catalog#_r=&collection=&country=54&dtype=&from=1991&page=1&ps=&sid=&sk=&sort_by=nation&sort_order=&to=2021&topic=&view=s&vk=
	UNICEF	No	Yes	Yes	Format: SPSS Satistics Data File Format (.sav) Must request access to available datasets, giving description of your research. Can only view or download once the application is accepted.	https://mics.unicef.org/surveys
	World Bank	Yes	Yes	Yes	Format: ASCII, SPSS + SAS Syntax; CSV; SPSS; STATA Accessible under three conditions: • Data available from external repositories accessible via third party link. • Public can access after accepting terms and conditions e.g. no attempt will be made to identify respondents. • Can only access after registering/signing in and agreeing to terms and conditions.	https://microdata.worldbank.org/index.php/catalog#_r=&collection=&country=54,140&dtype=&from=1890&page=1&ps=&sid=&sk=&sort_by=nation&sort_order=&to=2021&topic=&view=s&vk=

	World Health Organization	Yes	Yes	No		https://extranet.who.int/ncdsmicrodata/index.php/catalog#_r=&collection=&country=54&dtype=&from=1999&page=1&ps=&sid=&sk=&sort_by=nation&sort_order=&to=2019&topic=&view=s&vk=
State of Palestine	ILO Survey Catalogue	Yes	No	Yes	Log in is only open for ILO members	https://www.ilo.org/surveyLib/index.php/catalog?sort_by=rank&sort_order=desc&sk=palestime#_r=&collection=&country=83&dtype=&from=1975&page=2&ps=&sid=&sk=&sort_by=nation&sort_order=desc&to=2020&topic=&view=s&vk=
	FAO, United Nations	Yes	No	Yes		https://microdata.fao.org/index.php/catalog/1727/get-microdata
	UNICEF, MICS	No	No	No		https://mics.unicef.org/surveys
	GHDx	No, except labour force survey	Yes	Yes	Requires to log in, after that asks intended use and expected policy along with date of completion of process, For MICS, you need to registered as a user	http://ghdx.healthdata.org/geography/palestine
	World Bank	Yes	No except for Covid 19 survey	No		https://data.worldbank.org/country/PS
	IHSN	Yes	No	No		http://catalog.ihsn.org/catalog#_r=&collection=&country=83&dtype=&from=1890&page=1&ps=&sid=&sk=&sort_by=nation&sort_order=&to=2021&topic=&view=s

	OECD	No	Yes	No	Public access	https://data.oecd.org/searchresults/?hf=20&b=0&r=%2Bf%2Ftype%2Fdatasets&q=West+Bank+and+gaza+2010-+2021+household+surveys&s=desc(document_publicationdate)&l=en
	UNHCR	Yes	Yes	No		https://microdata.unhcr.org/index.php/catalog/central#_r=&collection=&country=83&dtype=&from=2011&page=1&ps=&sid=&sk=&sort_by=nation&sort_order=&to=2021&topic=&view=s&vk=
	IPUMS	No	Yes	Yes	Need to log in and provide reason for use and affiliated institution	https://international.ipums.org/international-action/sample_details/country/ps#tab_ps2017a
	Economic Research Forum	Yes	Yes	Yes	Need to request access by providing reason and purpose of data access, personal details and information of all candidates who will have access to the data	http://www.erfdataportal.com/index.php/catalog#_r=&collection=HHIES,LMPS,SYP,HLFS,HHHS,HEGS,Harmonized_Data&country=241&dtype=&from=2011&page=1&ps=&sid=&sk=&sort_by=nation&sort_order=&to=2021&topic=&view=s&vk=
The United Republic of Tanzania	Global Health Data Exchange	No	No	No	Microdata is not available for some surveys while for others it can be accessed through third party website after registration and logging in. However, some links to the third-party repositories do not work e.g. the link to access microdata for the Kigoma client experience survey	http://ghdx.healthdata.org/geography/united-republic-tanzania

	International Household Survey Network	Yes	Yes	No	Microdata is not available for most of the surveys while for some of the surveys it can be accessed via third party repositories on the links provided where you register and log in and access the data in DTA format after accepting the terms and conditions of microdata use	http://catalog.ihsn.org/catalog/8477
	World Health Organization	Yes	Yes	Yes	Microdata is available for PUFs and can be downloaded as CSV, DTA files after agreeing to the terms and condition. Access to licensed files is restricted and can be accessed after applying for access by filling the embedded application form. You can only download if your application is accepted.	https://extranet.who.int/ncdsmicrodata/index.php/catalog/history
	The World Bank	Yes	Yes	Yes	Submitting an application for access stating the intended use of microdata. Access to microdata for DHS and MIS surveys can be accessed via an external link that directs you to the IPUMS repository where you are required to register and log in. Microdata available as CSV, SPSS file and STATA files	https://microdata.worldbank.org/index.php/catalog/3991/get-microdata
	United Nations High Commissioner for Refugees (UNHCR)	Yes	Yes	Yes	Submitting an application for access stating the intended use of microdata. You can only download if your application is accepted.	https://microdata.unhcr.org/index.php/catalog/532/get-microdata
	Food and Agricture Organization of the United Nations (FAO)	Yes	Yes	Yes	Microdata accessible through third-party websites on the links provided where you are required to register and log in. Microdata is accessed after agreeing to the terms and conditions of use provide on the third-party website. Microdata available as ASCII+ SAS+STATA,CSV, STATA,SPSS files	https://www.fao.org/food-agriculture-microdata/en/

	International Labour Organization	Yes	Yes	Yes	Microdata for most surveys is available as aggregated variables for the whole sample. However microdata can also be accessed by contacting NBS director general on dg@nbs.go.tz	https://www.ilo.org/surveyLib/index.php/catalog/503
Vietnam	The World Bank	Yes	Yes	Yes	Varies based on survey however common formats included: SPSS, CSV, STATA, ASCII + SPSS + SAS	https://microdata.worldbank.org/index.php/catalog#_r=&collection=&country=196&dtype=&from=1890&page=1&ps=&sid=&sk=&sort_by=nation&sort_order=&to=2021&topic=&view=s&vk=
	IHME GHDx (Global Health Data Exchange)	Yes	Yes	Yes	CSV format	https://ghdx.healthdata.org/geography/viet-nam
	World Health Organization	Yes	Yes	Yes	Some microdata available - downloadable as DDI/XML or CSV format	https://extranet.who.int/ncdsmicrodata/index.php/catalog#_r=&collection=&country=196&dtype=&from=1999&page=1&ps=&sid=&sk=&sort_by=nation&sort_order=&to=2020&topic=&view=s&vk=
	UNICEF, MICS	No	Yes	Yes	Log in/account required to access raw microdata Final report with microdata included available for pdf download for all	https://mics.unicef.org/surveys
	Closer: The home of longitudinal research	Yes	No	No	Items are linked to external repositories.	http://atlas.closer.ac.uk/index.php/catalog#_r=&collection=&country=196&dtype=&from=1934&page=1&ps=&sid=&sk=&to=2021&topic=&view=s&vk=
	United Nations	Yes	Yes	Yes	Mix of open access data and request to access data.	https://www.wider.unu.edu/database/sur

	University UNU-WIDER				Open access mainly downloads in a .zip file. Once extracted the files are .dta (Stata files, but can be opened in R).	vey-data-growth-structural-transformation-and-rural-change-viet-nam-book
	IHSN International Household Survey Network	Yes	No	No	The microdata tab either provides a link to a third-party site to download the data or states that data access is not available. Although there is no clear information asking for someone to log in, I believe you may need to log in to access some of the data stored in the repository.	https://catalog.ihsn.org//catalog#_r=&collection=&country=196&dtype=&from=2011&page=1&ps=&sid=&sk=&sort_by=nation&sort_order=&to=2021&topic=&view=s
	Princeton University	No	No	No	All page's link to external repositories.	https://dss.princeton.edu/?utf8=%E2%9C%93&search_field=all_fields&q=vietnam
	FAOSTAT Food and Agriculture Organization of the United Nations	Yes	Yes	Yes	You can download a bulk of data on the main page, or indicator specific data. All data is in .CSV format	https://www.fao.org/faostat/en/#data
	Humanitarian Data Exchange (HDE) Data collated by UNHCR	Yes	Yes	Yes	All pages seem to provide open access for downloading. Depends on where the data originated. .CSV, .XLSX,. JSON, .PDF, .API	https://data.humdata.org/dataset/unhcr-population-data-for-vnm?force_layout=desktop
	The DHS Program - Demographic and Health Surveys	Yes	Yes	Yes	Login is required to download all datasets. STATA, SAS, Flat ASCII, SPSS, Hierarchical ASCII	https://dhsprogram.com/data/available-datasets.cfm

	Open Development Vietnam	Yes	Yes	Yes	All datasets seem to be available for download without having to log in or provide any information. Main download formats include: CSV, WMS, PDF, HTML, XLSX, DB_Table, ZIP, GEOJSON, XLS, XML Formats differ based on dataset topic.	https://data.vietnam.opendevelopmentmekong.net/dataset/?odm_spatial_range_list=vn&dataset_type=dataset