

**MCC Evaluation Microdata Documentation and De-Identification Guidelines**  
May 2017

## Contents

Scope.....	2
Background.....	2
Objectives .....	2
Prior to Data Collection .....	4
Storage and Transfer .....	5
Documentation and Dissemination – Evaluation Documents .....	6
Documentation and Dissemination – Microdata (Quantitative) .....	7
Documentation and Dissemination – Microdata (Qualitative) .....	9
Other Issues.....	10
Local Laws.....	10
Effective Date and Revisions to these Guidelines .....	10
Appendix 1: Terms – Microdata .....	11
Appendix 2: References and Other Reading Materials.....	12
Appendix 3: List of Annexes .....	13

# MCC Evaluation Microdata Documentation and De-Identification Guidelines

May 2017

## Scope

These [Evaluation Microdata Documentation and De-Identification Guidelines](#)<sup>1</sup> provide guidance to Millennium Challenge Corporation (MCC) staff and contractors, as well as the staff and contractors of partner governments that receive MCC funding on how to store, manage, and disseminate evaluation microdata collected as part of an MCC-funded program.

## Background

MCC collects microdata through surveys for a variety of purposes—from problem diagnostics, to project design, to evaluations. Unlike input and output data that is typically aggregated, microdata is unique in that it is collected—and can be reported—at the individual, household, enterprise, and/or community level. It also tends to have two distinguishing characteristics: it is personally identifiable and can be sensitive.

- **Personally Identifiable Information (PII)** is any information that can be used, on its own or in conjunction with other information that is linked or linkable to a specific individual, to determine the identity of an individual or otherwise locate or contact the individual. It includes:
  - *Direct Identifiers*: such as the individual’s full name, date of birth, mailing or home address, email address, telephone number, GPS coordinates, national identification number, physical/biological identifiers (physical appearance, through photo or video data collection, fingerprints, DNA, etc.); and
  - *Indirect Identifiers*: These include unique, observable or other characteristics that may enable re-identification even when direct identifiers are removed. Risk of re-identification is closely linked to the population the sample is drawn from and understanding how likely an outlier in the data is an outlier in the population<sup>2</sup>.
- **Sensitive data** is information that may pose a risk to the individual or firm if it is collected or released in a way that is linkable to the individual or firm responding to a survey. This type of data may include income, assets, or health status, the public release of which could harm survey respondents.

## Objectives

Microdata presents unique challenges to MCC using and disseminating the data. As an ethical matter, MCC, its agents, and its country partners must protect the survey respondents from harm that may be caused by public release of microdata. This obligation grows out of the *informed consent* process by which respondents are informed of the survey’s purpose and the identities of those who will have access to the collected information, and are promised that their responses will be kept confidential. Complying with this obligation is complicated by the fact that, unlike many federal agencies that collect PII as part of their programs, MCC does not have legal authority in its enabling statute to unconditionally protect the

---

<sup>1</sup> Version 3 (January 2017). Changes in this version: Included section on IRB and Data Collection; Clarified that MCC expects data to be de-identified, rather than anonymized; included recommendations for data encryption for storage and transfer; clarified required documentation and required data preparation/submission; provided clarification on qualitative data management; clarified DRB process.

<sup>2</sup> Indirect identifiers may include: Programmatic implementation identifiers. Depending on the program that is being evaluated, knowledge of who received the treatment and/or inclusion of a ‘treatment/control’ variable may significantly increase re-identification risk if the treatment population is well-known and greatly reduces the population size for which the survey is sampled; Assets, loan, and credit information: Some assets (land holdings, vehicles, livestock) are readily visible, and are potentially easy to use for re-identification. Multiple data points regarding loan and/or credit information may also allow for re-identification, such as including the name of the bank, date of loan, and value of loan. If there are not many lenders, or not many accessing loans, there may be a high risk of re-identification.

## MCC Evaluation Microdata Documentation and De-Identification Guidelines

May 2017

confidentiality of that information.<sup>3</sup> Moreover, the U.S. Government-wide statutes that commonly protect privacy do not apply to MCC’s evaluation-related microdata.<sup>4</sup> Without legal protection for the PII collected by MCC through its programs, such data could be subject to disclosure under the provisions of the Freedom of Information Act (FOIA),<sup>5</sup> which requires all US federal agencies to make their records available to persons who request them.<sup>6</sup> Another constraint to providing complete protection of survey respondent confidentiality is that MCC’s assistance to its country partners is provided through grants. This means that the microdata acquired with grant proceeds are assets and property of the grantee, which limits MCC’s ability to control access and dissemination and ultimately, its ability to ensure complete protection of the data. With these challenges in mind, MCC has defined three, often competing, objectives that guide its microdata protection principles and practices:

<b>Objective</b>		<b>Challenge</b>
<i>Protect privacy of survey respondents</i>	All data handlers (MCC, MCA, data collection firms, evaluation firms) should maintain promise of confidentiality made through informed consent process.	While it is clear how to minimize risk through removal of direct identifiers, assessing risk regarding inclusion of indirect identifiers, as well as assessing risk posed by other existing documentation that can link to this data, should be taken on a case-by-case basis.
<i>Facilitate verification of evaluation analysis</i>	Independent Evaluators (individuals or firms) are contracted to design, produce, and disseminate the results of the evaluations. MCC seeks to adhere to norms within the research community by ensuring that each evaluation’s analytic results are verifiable <sup>7</sup> - meaning MCC’s stakeholders, including policy-makers, researchers, implementers, and the general public - have the ability to analyze the same data as the Evaluators and replicate the statistical results.	Until now, many independent Evaluators use the complete, identifiable data for analysis. When de-identification is conducted AFTER analysis, data permutations - such as top/bottom coding, grouping, or even removing variables – may alter the data in a way that reduces a new user’s ability to verify original Evaluator analysis.  For <i>qualitative data</i> , the required de-identification process may severely reduce opportunity to verify results.
<i>Maximize usability of microdata</i>	MCC aims to maximize the usability of microdata for analysis above and beyond the original purposes of the evaluation, while respecting the terms of the informed consent. For example, the evaluation-related microdata could be used for analysis on other outcomes, or to examine differential impacts by geography, socio-economic, or gender characteristics. Such analysis may fall outside the original scope of the evaluation but have important development policy dimensions or business value.	When de-identification is conducted, data permutations - such as top/bottom coding, grouping, or even removing variables – may alter the data in a way that reduces its usability.  For <i>qualitative data</i> , the required de-identification process may severely reduce the usability of the data.

<sup>3</sup> See, for example the authorizing legislation for the National Center for Health Statistics (42 U.S.C. §242k, amended) and the United States Census Bureau (13 U.S.C §9, as amended). It should be noted that the majority of agencies whose authorizing legislation include such provisions are statistical agencies. MCC is a foreign assistance agency and not a statistical agency.

<sup>4</sup> For example, the Privacy Act of 1974, which protects privacy interests within the United States, does not apply because MCC’s programs focus on data collection of foreign citizens overseas. Similarly, the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA-Pub. L. 107-347, 116 Stat. 2899, 44 U.S.C. Sec 101), which provides broad legal privacy protections is unavailable to MCC because it is not a federal statistical agency and does not otherwise meet the requirements for applicability.

<sup>5</sup> The Freedom of Information Act (5 U.S.C. 552 (1996)) was first passed in 1967 and gives the public a statutory right of access to federal agency records with the aim of encouraging government accountability through transparency.

<sup>6</sup> There are exemptions from disclosure under FOIA that allow MCC to provide a measure of protection, thereby making the risk unlikely that microdata with PII would have to be released (for example, Exemption (b)(6) of FOIA exempts “personnel and medical files and similar files the disclosure of which would constitute a clearly unwarranted invasion of personal privacy” from disclosure under FOIA). However, the risk does exist and it limits MCC’s ability to guarantee respondent confidentiality fully.

<sup>7</sup> A secondary objective here is reproducibility – to enable the same study design and methods to be reproduced in a different context at a different point in time to see if the same results can be reproduced elsewhere.

# MCC Evaluation Microdata Documentation and De-Identification Guidelines

May 2017

## Prior to Data Collection

The protection of human subjects begins prior to data collection, with the Evaluation Design Report and survey materials, when the Evaluator *defines what data needs to be collected and why*. If PII does not need to be collected, then it should not be included in the questionnaire. If data that is being collected is already publicly available and not sensitive, then strict promises of confidentiality are not needed. All this should be discussed and agreed between the Evaluator, MCC, and partner country staff prior to data collection to ensure the research protocol and corresponding informed consent statement align with the requirements of the study. With this in mind, MCC has the following guidelines:

**Training.** To manage independent Evaluator contracts and oversee microdata collection efforts, all MCC Monitoring and Evaluation (M&E) staff should complete training on the protection of human subjects (to be renewed every 3-5 years). There is a free, online course available through the National Institute for Health - <https://phrp.nihtraining.com/users/login.php>. MCC recommends that all relevant MCA staff and evaluation contractors (Evaluators, data collection firms) also complete a similar training.

**Institutional Review Board (IRB) clearance.** As per MCC's standard independent Evaluator contracts, all independent Evaluators must submit a research protocol that covers the **entire evaluation period – whether by multiple submissions to the IRB or annual renewals - to a registered Institutional Review Board<sup>8</sup> (IRB)** to ensure appropriate study protocols are in place to protect the human subjects involved in evaluation-related microdata collection, storage, and dissemination. To do this, Evaluators must first assess local requirements, with the support of the MCA, and determine: (i) if there is a local IRB and if that local IRB is registered, and (ii) if any additional local requirements (such as cost) conform to international standards. If the local requirements do not conform to international standards, the Evaluator will work with MCC and MCA on a case-by-case basis to determine the way forward.

In MCC's experience, IRB costs should be estimated at \$3,000-\$5,000 depending on how many years the protocol must be in place. The process of submitting a protocol to an IRB can range from a few days to few months depending on the complexity of the study, the IRB schedule, and local requirements. The time and cost necessary for IRB review should be built into the Evaluator's work plan from the beginning.

**Informed Consent statement.** The MCC Disclosure Review Board (DRB) has approved a generic informed consent statement for *quantitative* data collection ([Annex 1](#)). Although this statement is recommended, MCC recognizes the final informed consent statement **must be reviewed and cleared by the Evaluator's IRB**. In instances where the IRB requires changes to the standard informed consent statement, Evaluators are requested to flag these changes during the MCC Evaluation Management Committee (EMC) review of survey materials to document any changes in how data will be collected, stored, and/or disseminated. The MCC M&E lead will determine what, if anything, needs to be flagged and discussed with the DRB.

---

<sup>8</sup> The IRB must be registered with the Office for Human Research Protections (OHRP) within the U.S. Department of Health and Human Services (HHS) and acceptable to MCC. While many countries have established country-level IRBs, not all do, and the quality of these IRBs varies across countries. For this reason, while the Evaluator must follow any country-specific requirements for the protection of survey respondents, the Evaluator must also submit the research protocol package to an international IRB.

# MCC Evaluation Microdata Documentation and De-Identification Guidelines

May 2017

## Storage and Transfer

While following the IRB approved research protocol, microdata handlers (data collection firms, Evaluators, MCA staff, MCC staff), should also ensure the following:

***Storage and Disposal of Paper Questionnaires.*** When used, paper-based questionnaires should be stored for two distinct functions:

- **Verification of survey responses.** This function should be addressed as soon as possible upon completion of data collection and alongside data entry. Double-data entry, with paper-based verification of discrepant responses, yields nearly perfect correspondence between responses and keypunched data and should be required of all *quantitative* surveys. Once this has been completed, the *non-respondent-identifier* sections of the questionnaire should be immediately and securely destroyed (shredded or burned depending on local resources).
- **Re-contact information.** When there is any need to maintain respondent identifier information – such as required for conducting a panel survey - this information should be linked to the remainder of the data through non-identifying unique codes, then removed from the remainder of the data, and securely stored. Securely, electronically scanned versions of these may be preferable to the originals. Once securely stored, any paper-based versions should be immediately and securely destroyed (shredded or burned depending on local resources).

***Microdata Storage.*** Once data collection ends, there should be specific practices in place to protect confidentiality of the microdata during storage. This includes actions by any data handlers, such as: encrypting data files; employing password protection on data systems and data encryption; and requiring relevant stakeholders to sign non-disclosure agreements. As per MCC information technology standards, the **end point encryption software should meet [AES-256 encryption standards or above](#)**. In MCC, all evaluation microdata should be stored in S:\DPE Data.

***Microdata Transfer/Access.*** When sharing identifiable data files – data with PII, either direct or indirect identifiers - data handlers should use a secure file transfer system and controlled access to the storage mechanism, considering the following:

- Ensure all communication channels are encrypted, especially Wi-Fi connections;
- File transfers should occur only through https connections;
- Use of hyperlinks for connections should be prohibited; instead, users should only connect to trusted sites by manually starting a new web-browsing session;
- As a last resort, password protect and encrypt all PDFs or other document types if there are no other solutions available for secure file transfers. Send passwords via a separate email or phone the recipient.

**MCC Evaluation Microdata Documentation and De-Identification Guidelines**  
May 2017

**Documentation and Dissemination – Evaluation Documents**

To facilitate access to and usability of evaluation microdata, **a sub-set of evaluation documentation deliverables will be posted on the MCC Evaluation Catalog and must be Section 508<sup>9</sup> compliant.** When necessary, this may require submitting an ‘internal only’ version of a document, as well as a ‘public-use’ version of the document (for example if the Evaluation Design Report contains geographic identifiers that may enable future re-identification of the respondent(s)). Table 1 summarizes the required documentation and format for documentation that must be made publicly available.

Document	Requested Format	Description
<i>Evaluation Design Report, Baseline Report, Interim Report(s), Final Report</i>	Word or searchable PDF	<p>These documents (deliverables required under MCC contracts) provide necessary design and analytical information for users of the data. <u>Evaluators should ensure that all public use documents/reports have been reviewed and edited to remove any references, such as geographic locations, that may threaten or undo data de-identification efforts.</u></p> <p>With regard to versions of Evaluation Design Reports (EDR), under current contracts MCC requires Evaluators to update the EDR as needed over the life of the evaluation. Any revisions should be documented in the EDR so that course corrections/revisions are clearly documented. In the event that one Evaluator inherits an evaluation from another, the original Evaluator’s EDR will be posted on the Evaluation Catalog along with the new Evaluator’s EDR.</p>
<i>Metadata File</i>	Nesstar (Annex 2) and PDF of Nesstar file	Once an EDR is cleared by MCC, the Evaluator should prepare the metadata file for the public evaluation catalog entry. The metadata can be updated/revised as necessary over the course of the evaluation. Please note: do not attach any data sets or related documents under the “other materials” or “external resources” sections.
<i>Informed Consent Statement</i>	Word, searchable PDF	The IRB approved informed consent statement should be published, either independently or as part of the questionnaire(s).
<i>Questionnaires (English and local language) and related documentation</i>	Original editable source and searchable PDF	<p>All survey questionnaires – baseline, interim, final - should be shared in a way that enables reuse by sharing the original editable source file. Evaluators may also submit a searchable PDF. Related documentation may also include sampling, field operations and interviewer manuals when needed for complete documentation of survey protocols. Any translation requirements should follow the Evaluator scope of work.</p> <p>For qualitative data, this documentation should include de-identified codebooks, field notes, researcher journals, etc. that would enable replication of the study.</p>

<sup>9</sup> In 1998, Congress amended the Rehabilitation Act of 1973 to require Federal agencies to make their electronic and information technology (EIT) accessible to people with disabilities. The law (29 U.S.C. § 794 (d)) applies to all Federal agencies when they develop, procure, maintain, or use electronic and information technology. Under Section 508, agencies must give disabled employees and members of the public access to information that is comparable to access available to others. The United States Access Board discusses the Section 508 law and its responsibility for developing accessibility standards for EIT to incorporate into regulations that govern Federal procurement practices. More information available online at <https://www.section508.gov/>

## Documentation and Dissemination – Microdata (Quantitative)

For every evaluation, the following should be considered:

1. **Consider and document de-identification strategy early.** De-identification efforts often require data permutations - such as suppression of specific variables' values, including, top and bottom coding, conversion of continuous variables to categorical or removal of any identifiable variation. Even if microdata does not need to be submitted to MCC until ALL data rounds are completed, the Evaluator should consider their de-identification strategy early and prior to analysis and document it in the DRB Data Package Worksheet for each round. Evaluators are encouraged to share their de-identification strategy with MCC as early as baseline to discuss implications for future verification of analysis and public and/or restricted access of microdata.
2. **For each data collection round, Evaluator should submit data documentation package.** Unless otherwise contractually required (i.e. a contract is expiring) or demanded by stakeholders (MCC, partner countries, other), the microdata files (the STATA files) do not need to be submitted to MCC for DRB review. However, upon completion of a data collection round (baseline, interim(s), final), the Evaluator should submit the survey materials (informed consent, questionnaire(s), updates to the metadata, etc) and the completed DRB Data Package Worksheet (Annex 3) to MCC as deliverables.
3. **Submit complete data files of ALL data rounds as one data package for DRB Review:** MCC aims for the microdata that is released as public and/or restricted-access to be as complete as possible. This means ALL data that was collected as part of the survey is included in the data package, not just constructed variables produced for the evaluation report. Unless otherwise agreed with MCC and stakeholders, Evaluators should plan to package ALL data rounds (baseline, interim(s), and final) as ONE data package for the DRB to review for public and/or restricted-access use. This is to ensure consistency in how de-identification of data is managed across data rounds, minimize risk of re-identification across rounds, and reduce costs. In cases where an Evaluator's contract will expire before the final round, or there is demonstrated demand for early rounds (baseline; interim(s)), then the Evaluator and MCC will discuss appropriate management of the microdata and work to ensure de-identification is managed in a way that considers publication of future rounds.
4. **Separate de-identification code from analysis code.** As a standard deliverable, MCC requests analysis code submitted as part of the final microdata package to enable/facilitate verification of evaluation analysis. This means the Evaluator should ensure any de-identification code is written separately from analysis code to ensure re-identification risks are minimized because de-identification code should NOT be publicly shared.
5. **Run analysis code on de-identified data.** When possible, Evaluators should run analysis code on the de-identified data files to demonstrate verification successes and/or challenges. This would improve documentation associated with reports and microdata, and complement the Transparency Statement (discussed below) to report what can, and cannot, be verified by the public-use and/or restricted-access data.
6. **Timely release of microdata.** MCC aims for microdata to be released in a timely manner to maximize usability, and therefore aims for release **no later than 6 months** following publication of the Final Report.

When ready to prepare microdata for public and/or restricted-access, **Evaluators should expect to submit the following package to MCC for DRB review:**

**MCC Evaluation Microdata Documentation and De-Identification Guidelines**  
May 2017

<b>Table 2: Summary of Quantitative Data Packages for MCC DRB</b>		
<b>Element</b>	<b>Requested Format</b>	<b>Description</b>
<i>DRB Data Package Worksheet</i>	Word ( <a href="#">Annex 3</a> )	Evaluators will follow this Worksheet, which draws on best practices outlined by the International Household Survey Network <sup>10</sup> , as well as recommendations of the Confidentiality and Data Access Committee (CDAC) <sup>11</sup> , a forum sponsored by the Office of Management and Budget’s Federal Committee on Statistical Methodology, the US Census Bureau, and USAID’s Demographic and Health Surveys.
<i>Data – Public Use and/or Restricted Access</i>	Stata 13 (or other format agreed with MCC)	This should be the complete data file – including the full dataset as collected and any constructed analysis variables. The ability to de-identify the data as per informed consent promises will inform whether or not this data is public use and/or restricted access. In some cases, <u>Evaluators have needed to produce public use files (to facilitate broad use of the data) AND restricted access files (to facilitate verification of results using data that cannot be made public).</u>
<i>Data Codebook– Public Use and/or Restricted Access</i>	PDF	Stata codebook output to review data - the codebook should include a label book as well as basic summary statistics including frequency and distribution information.
<i>Code– Public Use and/or Restricted Access</i>	Stata do file (or other format agreed with MCC)	This is the analysis code to produce the variables and analysis reported in the Evaluator report(s).
<i>Transparency Statement</i>	Searchable PDF	Evaluators should prepare Transparency Statement which states the extent to which data (public use and/or restricted access) can enable verification of results presented in the evaluation report. This would be discussed with the DRB and then finalized based on the final approved data file(s).

If necessary, this package should include any updates to the Metadata for the Evaluation Catalog.

The submission of the full data package for DRB review should be a [multi-step review process](#):

1. Evaluator and M&E Project Manager (PM) should agree on expected DRB review date as early as possible to confirm scheduling in line with Evaluator contract and work plan. This should be scheduled at least one month before the Evaluator’s contract expires given potential required follow-up after the DRB review.
2. Evaluator should submit full package to M&E PM. M&E PM should review Metadata and DRB Data Package Worksheet for clarity and completeness. This may require one round of revision based on the M&E PM requests for clarity and completeness.
3. Evaluator should submit full package to M&E PM. M&E PM and the M&E DRB members should establish a first-round review and provide feedback to the Evaluator on the proposed data de-identification process. This may require a second round of revision to the package based on feedback on documentation clarity and completeness, as well as proposed de-identification strategy.
4. Evaluator should submit full package to M&E PM for the confirmed DRB review date at least 2 weeks prior to the agreed DRB review date. Open Data Analyst will save in S:\DPE Data.
5. If any feedback/revisions are required following DRB review, Evaluator should revise and resubmit full package to M&E PM with documented responses to DRB feedback to ensure timely virtual review and clearance of the full package.

<sup>10</sup> <http://www.ihsn.org/home/node/137>; <http://www.ihsn.org/HOME/node/138>

<sup>11</sup> <http://fcs.m.sites.usa.gov/committees/cdac/>



**MCC Evaluation Microdata Documentation and De-Identification Guidelines**  
May 2017

**Documentation and Dissemination – Microdata (Qualitative)**

As of December 2016, MCC does not expect *qualitative* data to be prepared for public-use given unknowns<sup>12</sup> regarding de-identification and usability of qualitative data. In cases where the informed consent allows for qualitative data to be (i) made available to MCC and/or (ii) considered for restricted-access use, the Evaluator should prepare a full data package. For qualitative data that can be made available on a restricted-access basis, the data package will be held at MCC until the restricted-access mechanism is developed. At that point, the qualitative data package will be reviewed by the DRB and considered for restricted-access dissemination.

When preparing qualitative data for storage and/or future consideration of restricted-access, in addition to the documentation requested above in Table 1, **Evaluators should expect to submit the following package to MCC:**

<b>Table 3: Summary of Qualitative Data Packages</b>		
<b>Element</b>	<b>Requested Format</b>	<b>Description</b>
<i>DRB Data Package Worksheet</i>	Word <b>(Annex 3)</b>	To prepare microdata, Evaluators will complete the Worksheet, which draws on best practices outlined by the International Household Survey Network <sup>13</sup> , as well as recommendations of the Confidentiality and Data Access Committee (CDAC) <sup>14</sup> , a forum sponsored by the Office of Management and Budget’s Federal Committee on Statistical Methodology, the US Census Bureau, and USAID’s Demographic and Health Surveys.
<i>Data –Restricted Access (if allowable)</i>	Format agreed with MCC	For qualitative data, all direct identifiers should be removed. All indirect identifiers may remain in the dataset to the extent considered appropriate given the informed consent promises of confidentiality.
<i>Transparency Statement</i>	Searchable PDF	Evaluators should prepare a draft Transparency Statement which states the extent to which data (restricted access) can enable verification of results presented in the evaluation report. This would be a stand-alone document in the Evaluation Catalog, alongside the evaluation analysis report.

**If MCC cannot hold the identifiable data and/or the data cannot be considered for future restricted-access use, Evaluators must still submit the DRB Data Package Worksheet and Transparency Statement.**

<sup>12</sup> Evaluators and MCC are encouraged to watch the discussions here on transparency + qualitative research - <https://www.qualtd.net/>

<sup>13</sup> <http://www.ihsn.org/home/node/137>; <http://www.ihsn.org/HOME/node/138>

<sup>14</sup> <http://fcs.m.sites.usa.gov/committees/cdac/>

# MCC Evaluation Microdata Documentation and De-Identification Guidelines

May 2017

## Other Issues

### Guidelines for Evaluators Sharing Data prior to MCC DRB Review

1. The evaluation team is free to pursue analysis and publications beyond the scope of the evaluation *in accordance with the informed consent process, IRB approved research protocol, and the terms of the contract with MCC*. MCC will not cover the costs of such an analysis as it falls out of scope of the evaluation contract.
2. MCC expects the evaluation team to adhere to MCC's protection of human subjects requirements by submitting the proposed analysis protocol to their IRB, informing the IRB of the additional analysis that is outside of the original scope of the project, and adhering to the IRB's recommendations for data sharing.
3. The evaluation team should include in the submission to the IRB an assessment of risk associated with providing the summary statistics/data for re-identification of future public-use and/or restricted-access data available on the MCC Evaluation Catalog. Will provision of the summary statistics/data enable/increase re-identification of that data?
4. The evaluation team should inform MCC of the IRB decision for this specific request.
5. MCC requests the evaluation team inform MCC about the results of this additional analysis—i.e. work with us to circulate the initial findings to interested stakeholders. If/when any manuscript is published, the evaluation team should share it with MCC to circulate to interested parties.
6. When the evaluation team prepares the public-use and/or restricted-access data for the DRB, they should include a discussion of the IRB decision to share summary statistics/data outside the original evaluation team to inform the DRB's decisions around release of public-use and/or restricted-access data.

## Local Laws

In addition to the guidelines found here, data handlers should consult with the MCA to comply with all relevant local laws, including those that may supersede or conflict with these guidelines or prevent dissemination. Where local law prevents compliance with these guidelines, the MCC and partner country M&E and legal staff should prepare a memo to guide staff and contractors on how to proceed.

## Effective Date and Revisions to these Guidelines

These guidelines are effective as of January 23, 2017. These guidelines may be revised and updated from time to time, and such revision will be promptly posted on the MCC website. If the guidelines are updated during the course of one evaluation or contract, staff and contractors are requested to apply the most recent, approved version to their work to the extent possible.

**MCC Evaluation Microdata Documentation and De-Identification Guidelines**  
May 2017

## Appendix 1: Terms – Microdata

Data Type	Description	Rules
Raw data	This is the data directly collected by the data collection firm and submitted to the Evaluator.	<p>Raw data that is otherwise publicly available, low risk, and/or contains no PII should be submitted to MCC as a deliverable and packaged for public use.</p> <p>Raw data that includes sensitive, PII data is USUALLY NOT submitted to MCC. Management of this data will be discussed on a case-by-case basis until MCC has an available option for rigorous management of identifiable data. This is particularly crucial for the continuation of a study when there is a change in Evaluator.</p>
Public-Use data and code	Public use data is de-identified to be in line with promises made through the informed consent. Public use data should include the <u>full set of variables as collected, as well as any constructed variables for analysis, as well as the code used to construct those variables and conduct analysis.</u>	<p>Evaluators should submit the code for constructing the analysis variables to facilitate verification of results. Simply submitting the analysis file of constructed variables is insufficient as it does not allow for re-tracing how variables are constructed.</p> <p>MCC aims for a public-use file to be prepared and disseminated for EVERY evaluation, if only to meet our ‘maximize usability’ objective. If required data permutations for de-identification limit ability to verify results and/or maximize usability, then a restricted-access file must be considered.</p>
Restricted-Access data and code	This is data that removes direct identifiers (names, addresses, GPS coordinates), but retains indirect identifiers. As with public-use data, restricted-access data should include the <u>full set of variables as collected, as well as any constructed variables for analysis, including the code used to construct those variables and conduct analysis.</u>	<p>When a public-use file cannot meet verification and/or usability objectives, and the informed consent allows for restricted-access to the data, then Evaluators will be requested to prepare restricted access files <b>in addition to</b> or <b>instead of</b> the public-use file (as applicable).</p> <p>For restricted-access files still considered ‘high risk’ by the DRB, there is currently no mechanism for dissemination. Evaluators are asked to prepare these files and either (i) store them until MCC establishes the mechanism or (ii) submit to MCC to store if it is not possible for the firm to store.</p>

**MCC Evaluation Microdata Documentation and De-Identification Guidelines**  
May 2017

## Appendix 2: References and Other Reading Materials

Alderman, Harold, Jishnu Das, and Vijayendra Rao. 2016. "Conducting Ethical Economic Research: Complications from the Field." In *The Oxford Handbook of Professional Economic Ethics* edited by George DeMartino and Deirdre McCloskey. Oxford University Press, April.

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

NISTIR 8053, De-Identification of Personal Data, Simson Garfinkel, September 2015, National Institute of Standards and Technology, Gaithersburg, MD. <http://dx.doi.org/10.6028/NIST.IR.8053>.

NIST 2016, De-Identification of Government Data <http://csrc.nist.gov/publications/PubsDrafts.html#SP-800-188>

Glennerster, Rachel, and Shawn Powers. 2016. "Assessing Risk and Benefit: Ethical Considerations for Running Randomized Evaluations, Especially in Developing Countries." In *The Oxford Handbook of Professional Economic Ethics* edited by George DeMartino and Deirdre McCloskey. Oxford University Press, April.

Hanson, Heather and Catherine Marschner. January 2015. "Transparency" Millennium Challenge Corporation Principles into Practice. <https://assets.mcc.gov/reports/paper-2015001163301-principles-transparency.pdf>

Miguel, E., C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, et al. 2014. "Promoting Transparency in Social Science Research." *Science* 343 (6166): 30–31. doi:10.1126/science.1245317.

Poe, Ted. (10/20/2015). "H.R.3766 - Foreign Aid Transparency and Accountability Act of 2016." Legislation. July 15. <https://www.congress.gov/bill/114th-congress/house-bill/3766>.

Ryan, Paul. 2016. "H.R.1831 - 114th Congress (2015-2016): Evidence-Based Policymaking Commission Act of 2016." Legislation. March 30. <https://www.congress.gov/bill/114th-congress/house-bill/1831>.

Sturdy, Jennifer, Sixto Aquino, and Jack Molyneaux. 2014. "Learning from Evaluation at the Millennium Challenge Corporation". *Journal of Development Effectiveness*. Taylor & Francis. DOI:10.1080/19439342.2014.975424

## Appendix 3: List of Annexes

Annex 1: [Informed Consent Template \(Quantitative\)](#)

Annex 2: [Metadata Template](#) (tutorial and template) (internal MCC document)

Annex 3: [DRB Data Package Worksheet](#)