This document provides guidelines for preparing Terms of Reference (TORs) for consultants and consulting firms, where the purpose of the engagement includes data collection activities. These guidelines are not mandatory, but are provided as a guidance. Quite often staff engage consulting firms for projects which may include data collection tasks, such as: surveys, impact evaluations, physical measurement, or for geo-spatial data.

The guidelines do not prescribe how the overall TORs are to be drafted, but they draw heavily on TORs prepared by a cross-section of Task Team Leaders (TTLs), who have used them successfully in their projects. The main emphasis of these guidelines is on the data related aspects such as,

- Data - what dataset(s) are being delivered
- Data Format - specifications of format(s) in which data is to be prepared for delivery
- Metadata - documentation about the dataset(s) that helps understand the data, its structure, how it was prepared and how it is to be used
- Data confidentiality, ownership, public availability, and re-use
- Ethical Provisions in data collection, where applicable.

These aspects of the data are important in terms of dataset cataloguing, curation and future use by others in or outside the World Bank. The guidelines can be used while drafting: TORs for hiring consulting firms to do data collection; ‘Request for Proposal (RFP), or Request for Expression of Interest(REOI) from vendors; or Master Agreements with data subscription vendors, who collect, for example, specialized geo-spatial data time series (e.g. on climate or on renewable energy).

Given below is the suggested structure of the TORs. The TORs template that follows, describes the various sections in it.

1. Background
2. Objectives
3. Scope of Work
4. Approach and Methodology
5. Duties and Responsibilities
6. Deliverables and Timeline
7. Data Deliverables
8. Consultant Qualifications
9. Data Confidentiality/Data Ownership/Data Reuse
10. Ethical Provisions
11. Budget, Contract Management and Payment Schedule
12. Selection Criteria
13. Training by WBG for Data Collection (Optional)
14. Specific Inputs from WBG (Optional)
15. Pricing (optional)
16. Potential for Downstream Work (Optional)
17. Annexes (if required)
Terms of Reference (TORs) for Data Collection

TEMPLATE

The paragraphs below describe the sections in a Terms of Reference document that are suggested for projects which include data collection activities as the principal tasks in a project, or in some cases, are undertaken to support other major tasks in a project. This template may be used, not only for Bank-executed projects where TTLs develop the ToRs, but also for Recipient-executed ones where TTLs provide advice on ToR drafting to country counterparts. This document provides guidance mainly for the data deliverables, data documentation (metadata) and on the data ownership protections for the World Bank. TTLs may change the structure and content as appropriate.

1. Background

The opening section of the TORs offers an orientation to the reader about the overall project/program. It provides the context in which the tasks described are to be performed. It should also provide, in a brief paragraph, what data collection tasks are to be performed and what are the main expected outcomes. Depending on the complexity of the project/program, this section can be a few paragraphs or can be a couple of pages long.

2. Objectives

This section maps out the specific objectives of the engagement. Although the general rationale may have been introduced in the Background section, this section would spell out the additional details in terms of the evaluation or research question(s) being answered, or the objectives of an ongoing program of data collection on, for example, household or enterprise surveys.

The objectives section can be short or long, depending on the specific data collection project. But it is generally drafted in a clear and outcome-focused language.

3. Scope of Work

The scope of work describes the parameters that limit the work to be performed. That is, besides describing the work it also sets boundaries around it. The scope would,

- Describe broadly the data collection work to be performed
- Set limits on:
  - Actual time-period estimated for data collection,
  - Elapsed period over which these tasks should be performed,
  - Geographical boundaries over which data is collected,
  - Target groups for the data collection, and
  - Outputs/services etc.

If the data collection builds on an earlier effort or complements another, then the scope text should mention that. Also, it is useful to indicate issues, tasks, or other parameters that fall outside the scope of these data collection TORs.
4. Approach and Methodology

This section is especially necessary for monitoring and impact evaluation studies that require baseline, mid-line or end-line (follow-up) data collection. In such cases, the approach and the precise methodology for data collection can be a challenging part of the TORs. Key elements generally highlighted in this section include:

- The overarching methodological framework (for example, case study, sample survey, mixed methods, and so forth)
- Expected data collection and analysis methods with descriptions of any special tools (software) used to collect data.
- Outcome and output indicators that are being proposed, or any baseline data, target or counterfactuals that are proposed.
- Availability of other relevant data, such as existing local, regional, or national data or data from similar programs
- The process for verifying findings with key stakeholders
- Meetings or consultations or other interactions expected with stakeholder groups

5. Duties and Responsibilities

This section includes a specific list of tasks to be performed by the consulting firm/vendor for which it will be responsible. If the tasks involve multiple parties, then the responsibilities for each should be stated explicitly. The areas generally covered are: staff requirements, recruitment and hiring, training, supervision, questionnaire development, pilot testing, technical tasks such as interviewing, data quality controls, organization and logistics, workshops, and so forth.

Staff have the responsibility to review and ensure that all deliverables are consistent with the requirements in these TORs and are satisfactory.

6. Deliverables and Timeline

Expected outputs from the detailed tasks and responsibilities should be listed with a brief description of the output/deliverable and the planned timeline for each. Deliverables, such as -- the project inception report and work plan, intermediate reports on activities described in the ‘Duties and Responsibilities’ section, progress reports and final report, meetings/consultations/seminars, etc. should be listed in this section. Any data related deliverables should be listed separately along with the specifications. (See section below on Data Deliverables)

7. Data Deliverables

---

1 For ongoing survey data collection programs, or where the methodology is standard and well known, these details may be excluded at the TTL’s discretion. The methodological details can also be placed as an Annex to the TORs. Some TORs include the methodology under the ‘Duties and Responsibilities’ section

This is a separate section that highlights the data deliverables in a project. There are a few important deliverables about the data collected, which should be included in this section. These are the Data and the Documentation about the data or the Metadata.

Given below are guidelines for how the data deliverable may be specified. Depending on the data collected, TTLs should select from the data types specified, and add or modify the specification below as needed.

7.1 Data

Data Delivery: Data includes raw datasets, as well as, cleaned and validated datasets collected or procured under these ToRs, and used in producing final deliverable outputs, such as, but not limited to: reports, tables, and visualizations. In case of surveys, the raw datasets must include actual microdata from respondents. Also, raw data must be delivered regularly throughout the data collection process (e.g. weekly or monthly, depending on the duration of the exercise). This is critical to identify data quality issues (such as missing observations, high attrition rates, high rates of refusal) whilst the data collection is ongoing, so that they may be corrected in real-time.

Machine-readable format. Datasets should be provided in electronic form, or in files which are in a machine-readable format.

Data Security while delivering data. Appropriate security protocols must be observed in the transfer or transmission of datasets to the Bank. The suggested mechanism for a secure transfer of datasets between the consulting firm and the World Bank would be to use the Bank’s cloud-based OneDrive storage (5 Terabytes of storage per account), and which also provides a secure SSL/TLS connection and encryption of the dataset.

Bank TTLs would be responsible for: (i) creating a dedicated OneDrive repository for the data, and, (ii) Ensuring that only authorized collaborators (in the consulting firm and the Bank) are given access to the OneDrive repository.

The consulting firm would be responsible for ensuring that all data generated under this project (contract) is deposited in the dedicated OneDrive repository.

Data Format: should be selected based on the data type (survey, geo-spatial, sensor, transit, etc.), Staff should choose the appropriate format from the choices below in which data should be delivered.

i. Survey dataset(s) may be provided in STATA, SPSS, CSV, Excel, ASCII, etc. or any other suitable open format.

---

3 TTLs should note that some countries may have restrictions on raw or validated datasets, which contain sensitive and/or personally identifiable information (PII), to be transferred to the Bank, based on the countries’ national laws/regulations or Ethics provisions for PII or other sensitive data. If access to the raw datasets or any related data documentation is essential, then TTLs must negotiate with the country counterparts and the consulting firm, terms for such data transfer or access, and include them in the ToRs or in the Terms and Conditions of the contract to be issued by the Procurement Unit.

4 Transport Layer Security (TLS) or the Secure Sockets Layer (SSL) are cryptographic protocols that provide communications security over a computer network.
ii. Geo-spatial dataset(s) may be provided in open format shape files which can be opened by a GIS application of choice. This data could also be provided in any other readily available formats such as CSV, XLS (97 and later), XLSX (2007 and later), TSV, or any other suitable open geo-spatial data format. Preferred formats are shapefiles for Vector\(^5\) data, and GeoTiff for Raster\(^6\) data. Geo-spatial dataset(s) can also be collected for the following categories. Although these dataset(s) can also be provided in the formats above, it is recommended that data collectors should provide these data in open formats using the encoding standards given below for each.

- Hydrology – Hydrological time series data including ratings, gauges and sections. Encoding standard to be used: WaterML.
- Geology – Geological data. Encoding standard to be used: GeoSciML.
- Internet of Things (IoT) – Sensors and observations from IoT devices. Encoding standard to be used: Sensor Things API.
- Sensor Data – Time series data from sensors and the ground or mounted to platforms. Encoding standard to be used: O&M and SWECommon.
- Transit Data – Data published by transit agencies. Includes schedules of public transportation (trips, stops, times, routes, etc.). Encoding standard to be used: GML and GTS.
- Open Street Map (OSM) – Data from Open Street Map. Encoding standard to be used: GML and OSM.

See [http://www.opengeospatial.org/standards](http://www.opengeospatial.org/standards) for a list of acceptable formats.

### 7.2 Data Documentation (Metadata)

Metadata (documentation about the data). Documentation details vary depending on the type of data. The metadata below comprise documentation that is intended to help understand how the data have been created, what they are measuring, or in the case of geo-spatial data what they represent, and how they are to be used.

- **Survey dataset(s)**: Metadata necessary to understand the survey design, methodology, and the datasets must be provided. Data documentation provided must comply with international metadata standards. The most common standards are the Data Documentation Initiative (DDI) (Codebook version) and the Dublin Core standards, which are supported in the micro-data repository of the World Bank. All datasets must be documented in compliance with these

---

\(^5\) Vector Data uses the simple geometric objects of points, lines, and areas (polygons) to represent spatial features

\(^6\) Raster Data uses a grid to represent its geographic information. Points are represented by single cells, lines by sequences of neighboring cells, and areas by collection of grouping cells. Types of raster data include: satellite imagery, remotely sensed satellite data.

\(^7\) The list of metadata and other documentation in this Section is a suggested list. Depending on the data type – Survey, Geo-spatial, Physical measurement, or other – TTLs should specify more details as required
standards, which should foster compliance with the replication standard and guarantee the ease of use, quality control, and cataloguing of the dataset. Given below is a list of metadata documents.

i. Survey questionnaire in English and local languages which includes any/all changes from piloting;
ii. Questionnaire design;
iii. Interviewer manual;
iv. Completed household/community questionnaires for all households/villages interviewed;
v. Concise, but complete description of the sample design;
vi. Sampling weights to be used in data analysis;
vii. Description of the basic field procedures and quality control techniques;
viii. Guidelines for using the data, including, but not limited to,
   a. Uniquely identifying observations, including the identification variable to be used,
   b. how to link the various parts of the survey,
   c. codes for any items not pre-coded in the questionnaire, and
   d. problems encountered in the data and the solutions taken;
ix. Documentation of constructed data sets that will be available in the public use files; e.g. price indices, and aggregates of household consumption or income;
x. Descriptions of the files, including contents and names of the data files mapped to the corresponding sections in the questionnaire; and,
xii. System of variable names and labels, data descriptions and dictionaries.

• Geo-spatial dataset(s): Given below is a suggested list of metadata.

i. Title – name by which the dataset is known;
ii. Description – brief summary including the purpose for which the dataset was developed;
iii. Theme – common use word or phrase used to describe the subject of the dataset;
iv. Place – geographic name of the location covered by the dataset;
v. Data presentation format – this refers to the Vector or Raster formats used to represent the geo-spatial data, including any specific standard file format used;
vi. Map projection – name of the map projection used;
vii. Bounding Coordinates:
   a. Western-most coordinate of the limit of coverage expressed in longitude
   b. Eastern-most coordinate of the limit of coverage expressed in longitude
   c. Northern-most coordinate of the limit of coverage expressed in latitude
   d. Southern-most coordinate of the limit of coverage expressed in latitude.
7.3 Data Entry Software

The data entry and/or data validation software, as well as, the hardware (tablets/phones), if procured by the consulting firm for this project with funds provided under the contract, must be handed over to the TTL when the project’s final report is delivered.

7.4 Data Quality

This section specifies the quality assurance aspect of data collected at a unit-level in a survey. This also becomes necessary to specify contractual consequences for fraud or negligence in collecting survey data. Given below are sample paragraphs that cover these two important aspects. TTLs may amend these depending on their specific data collection situations.

“The contracted firm’s survey manager and fieldwork supervisors will check all completed questionnaires and will conduct a minimum of 10 percent of spot checks to verify the accuracy of the data recorded and, where necessary, clarify with respondents any inconsistencies in their answers with that of the enumerator. The survey manager and fieldwork supervisors will randomly check interviewers and accompany them on some interviews. Representatives of the World Bank may also accompany some survey teams during pre-testing/piloting as well as during the survey to ensure quality, as well as provide on-site supervision of the data entry and database cleaning processes as a means of additional quality control.”

“The contracted survey firm will be considered to have failed to comply with this contract if, based on a random and representative sample, it is determined that either (i) it is shown that 1% or more of the questionnaires that are presented were filled without the Survey Firm having visited the respondent, and (ii) it is shown that 1% or more of the questionnaire is inconsistently completed. In either of these situations, the World Bank will reserve the right to request a repetition of the work or the option of not paying for the work done (being reimbursed for any initial payment)”

7.5 Names, Locations, Mobile Contacts of Respondents

TTLs, in some cases, may want to have the contact information of survey respondents (names, locations or mobile numbers) to contact them for verifying/validating their responses. Such data should normally be considered as a deliverable under the “Data Deliverables”.

8. Consultant Qualifications

---

8 TTLs should note the following and include in the ToRs as appropriate. For conducting surveys and collecting data, the consulting firm could use the World Bank’s Survey Solutions (available free of cost), CS-Pro, CWEST, SurveyBe, CASES, Open Data Kit (ODK) or any other suitable commercially available Computer–Assisted Personal Interviewing (CAPI) software available on mobile devices (tablets/phones).

9 An exception can occur in countries where ethical considerations and/or national laws/regulations disallow the data collection consulting firm from providing such personally identifiable information to the World Bank.
This section lays out the mix of knowledge and skills needed to execute the scope and methodology described in the TORs. TTLs may describe the profile of the specialists or the team needed for the tasks, including the technical competence in the sector or discipline (statistical surveys, physical measurements, geo-spatial, or other data collection), language proficiency and in-country or regional experience.

9. Data Confidentiality/Data Ownership/Data Re-use

Data confidentiality and data ownership are important aspects of the data collection activities and should be explicitly stated in the TOR. The following text, or a variation thereof, that preserves the confidentiality, data ownership, and data re-use aspects should be included in the TORs.

All data collected under these TORs including the documentation about the data and its collection methodology are confidential and are the property of the World Bank. The consulting firm executing these TORs, must protect the confidentiality of responding establishments and individuals participating in the data collection at all stages. The consulting firm must ensure that data or the documentation referred to above, collected or compiled under these TORs, are not distributed for commercial or non-commercial purposes to third parties, nor will they be used by the firm, their staff and/or consultants, or their sub-contractors for purposes other than those expressly stated in these TORs, without the written approval of the World Bank prior to, during, or after completion of the assignments.

10. Ethical Provisions

Data collection situations often warrant compliance with ethical standards of social science research that involves human subjects. While preparing, and testing the survey questions and the surveying process itself, the consulting firm must comply with such standards. An informed consent by the survey respondent to participate in the interview is suggested. Efforts should be made by the surveyor to balance transparency of the survey and confidentiality of the respondent. Ethical provisions for data collection in use by the national statistical offices or ministries of the government should be adhered to, unless they do not exist or are inadequate. In such situations, the consulting firm should be encouraged to comply with the Institutional Review Board (IRB) in the country, or with other regional or international review boards.

The link below is to one such IRB.

Western Institutional Review Board, WA, USA

Annex 1 gives some guidelines to researchers and evaluators on the Ethical Provisions in data collection.

---

10 Based on ethical considerations and/or on national laws/regulations, countries may not allow Bank ownership of sensitive or personally identifiable data (e.g. health related data of individuals). In such situations, prior to issuance of the Purchase Order, TTLs with the help of the Corporate Procurement Unit, should negotiate access to the data for specific Bank staff who would be working on such sensitive data and include such exceptions in the section on Data Confidentiality/Data Ownership/Data re-use.
Both, the survey transparency and the respondent’s confidentiality can be compromised in certain situations where a respondent may be answering personal questions in the confines of his/her home, but in the presence of family or community members. While for benign information, such as, on demographic characteristics, or consumption patterns, this is less of an issue, for information on medical conditions (HIV/AIDS), or on issues of domestic violence, or other sensitive matters, the survey, could turn out to be a traumatic experience for the respondent.

11. Budget, Contract Management and Payment Schedule

For TORs where the financial details have been worked out, this section includes the budget allocated and the payment schedule based on the delivery of outputs. This section may also include names and titles of individuals managing the contract, as well as the start and end dates for the project. These details are typically included after contract negotiations have been completed. Please note that names and titles of individuals should not be included in the TOR during the REOI/RFP stage.

12. Selection Criteria

Where the selection of the consulting firm is being done through a competitive process, this section is required. The detailed technical evaluation criteria and the relative weightage for each should be outlined here. When finalizing the TOR for inclusion in the contract after the selection has been completed, selection details relevant to the RFP phase should be removed or revised to be relevant to the contract implementation phase.

[Sections 13 through 16 are optional]

13. Training by World Bank staff for Data Collection

This section is optional. Depending on the specific data collection project, World Bank staff may include in the TORs this section to provide training for the field staff in questionnaire development for surveys and in pilot testing of the survey on a selected sample. This section should provide more details on the training to be provided and the timeline for it.

14. Specific Inputs from World Bank

This section is optional. Depending on the data collection project, the TTL may agree to provide certain specific services or inputs as part of the project. This may include training (#13 above), technical assistance/advice, or other facilities. It is important to include these inputs in the TOR as this may have a significant impact on the methodology and approach as well as the price proposed by the firms.

15. Potential for Downstream Work

This section is optional. It can be relevant when the TORs are being drafted for ‘Request for Proposal (RFP)’ or for ‘Request for Expression of Interest’, in which a longitudinal/panel survey is being undertaken with plans for follow-up surveys as the 2nd or 3rd phases of the project. In such cases, it is likely that the TTL may prefer to continue with the same consulting firm that worked on the initial survey. Such a section makes it easier to contract the same firm based on the ‘Single Source Selection’ justification in the 2nd or 3rd phase of the project. This section may also be relevant for cases where baseline, mid-line and
follow-up surveys are planned in a project, where the same firm with the project knowledge may be appropriate.

16. Annexes

This section is optional. Annexes may be included to provide more details of the ‘Approach and Methodology’ described in the TORs, or any other section that needs more elaboration.
ANNEX C.1: Ethical Provisions in Data Collection for Research/Analysis/Evaluation

These guidelines on ethical provisions\(^\text{11}\) may be included in the ToRs for Bank-contracted firms hired for data collection. They are applicable to data collection tasks on sensitive topics or those involving human subjects. The purpose is to encourage and ensure that contracted firms adhere to the highest standards of ethics in data collection.

**Ethical Provisions**

1. Researchers and evaluators would be responsible for identifying the need for and securing any necessary ethics approval for the study they are undertaking. Technical proposals should include a description of the protocol for human subjects and a plan for securing ethical clearance in the technical proposal. Such a protocol may be from national or local ethics committees in countries in which the study will be undertaken, or other stakeholder institutions with formal ethics approval systems.

2. Researchers and evaluators should avoid harm to participants in studies. They should ensure that the basic human rights of individuals and groups with whom they interact are protected. This is particularly important with regard to vulnerable people. The wellbeing of researchers/evaluators working in the field should also be considered.

3. Participation in research and evaluation should be voluntary and free from external pressure. Information should not be withheld from prospective participants that might affect their willingness to participate. All participants should have a right to withdraw from research/evaluation and withdraw any data concerning them at any point without fear of penalty.

4. Researchers and evaluators should ensure confidentiality of information, privacy and anonymity of study participants. They should communicate clearly with prospective participants any limits to confidentiality.

5. Researchers and evaluators should operate in accordance with local laws and international human rights conventions and covenants. They should respect cultural sensitivities, such as: differences in culture, local behavior and norms, religious beliefs and practices, sexual orientation, gender roles, disability, age and ethnicity and other social differences.

\(^{11}\) These provisions are adapted from the Ethics Principles for Research and Evaluation established by the United Kingdom’s Department for International Development (DFID) and adopted by the Strategic Impact Evaluation Fund (SIEF) of the World Bank.