Annex A: Functional Requirements for Election Data Management System

DATABASE MODULES AND REQUIREMENTS

A complete database instance would allow for the following:
- Multiple (often simultaneous) events per instance (i.e. over the same time period, data can be captured simultaneously from multiple discrete forms attached to discrete lists of participants)
- Multiple (often simultaneous) forms per event (i.e. over the same time period)
- Multiple (often simultaneous) reporting periods per form (i.e. forms can be activated for multiple (non-synchronous) events using either the same or discrete lists of participants)

This model has two implications:
- Participants, including system users and their related user permissions are linked at the form level (see below on **access control lists**)
- Each form must include a unique prefix code to be used when reporting data for that form, as must each form that is to be completed synchronously by discrete lists of participants so that the data can be processed and summarized separately

Database Modules, in order of priority

- Set-up, configuration and management of geography, people, forms and events
- Data entry
- Data reporting
- Data quality
- Data summary

The code to support each of the core modular fashion such that each component could be operated in a stand-alone manner or in conjunction with another system.

SET-UP, CONFIGURATION AND MANAGEMENT OF GEOGRAPHY, PEOPLE, FORMS AND EVENTS MODULE

(Desired but not required) The system should include a configuration platform that allows end-users to specify the relevant administrative divisions/geographic structures for the observation effort as well as their hierarchy (e.g. country is the parent for state is the parent for county), create forms, create location lists with unique IDs, create participant lists with unique IDs and specify the location ID to which each participant is linked, specify system users, and set the time periods for events.

Data Export Options

A landing page for imports and exports should be provided under set-up and configuration. The data management system should be able to export datasets in a common data format (guidance on the format to be provided upon request), including but not limited to:

1. (Required) Functionality to export participant lists - members of the observation structure as well as system users such as data clerks or NDI staff - with relevant demographic data in CSV or XLS.
2. (Required) Functionality to export administrative structures/geographic hierarchy, which can include strata.
3. (Desired but not required) Possibility to export forms in an XLS format.

Data Import Options

(Desired but not required) A landing page for imports and exports should be provided under set-up and configuration. The data management system should be able to import the following datasets in CSV or XLS format:

- Administrative structures/geographic hierarchies
- Participants
- Locations
- Samples and Sub-samples
- Possibility to import forms in an XLS format.

Observer Management

The database should be able to manage contacts at all levels of the observation structure, including observers, supervisors and coordinators. Contact management records should include: name, primary telephone number, secondary telephone number, role/position within the organization (i.e. observer, supervisor, coordinator), geography and polling station assignment (sometimes called a location ID) as well as any polling station changes/updates. The system should be able to do bulk communication with any or all of the contacts in the system, typically via SMS.

> *Access Control Lists (ACLs):* The database should also include functionalities to configure access via role-based control lists with different levels of access available to data clerks, data managers, analysts, supervisors and observers.

DATA ENTRY MODULE

The database will collect all observation data. This may include: tables for the official polling station list; geographical structure of the country; observer contact lists; observed process and results data; and critical incident information. Additional requirements can include: verifying mirrored information from multiple observers; registration of phone numbers; and closed system management.

Data types that the system should be able to support are as follows:
- Boolean (yes/no)
- Numerical (i.e. results, number of registered voters)

- Categorical or interval (i.e. number of polling station officials with the options being, for example: zero, between one and five, five, and more than five if five is the expected number of officials in any given polling station)
- Multiple selections (i.e. tick one or more answers given multiple choices)
- Multiple answer (i.e. tick one and only one answer given multiple choices)
- Short text strings sent via SMS
- Long-answer responses/open text fields to enable data operators to manually provide additional explanations of responses submitted by the observers. Each question should have an option to add or toggle on such an open text field.
- Control - matching
- (Desired but not required) Multimedia attachments such as photos and videos including photos of results sheets posted at polling stations as well as observer arrival selfies

Data entry should be possible via the following options:

1. Manual Data Collection through a GUI Interface - Observation information can be entered manually (by phone operators or entry clerks) in a browser. For each reporting period (or, in the case of SMS reporting, for each text message), data operators are able to filter polling station records by administrative divisions, supervisor, sub-samples (if applicable), polling station type (e.g. capital, urban or rural), and reporting period "status:" a complete report, an incomplete report (i.e. a report with some question responses missing), a problematic report (i.e. a report that has triggered a data integrity check), and a missing report (i.e. no question responses have been submitted). The status filter enables operators to rapidly identify those observers whom they need to contact in order to get complete, internally coherent data for each reporting period. Operators will also need to be able to perform searches to look up the relevant form by its unique polling station code, or other (geographic) identifier.

2. Data Collection via SMS: The system should have the capability to automatically parse and enter a text message from an observer into the central database and present the summary report through an intuitive user interface. The processing system will also produce and send automatic responses to phone number that initiated the message, including: 1) confirmation of message receipt, 2) identification of problematic messages (i.e. messages that did not use appropriate formatting or that submitted an inappropriate response type for a given question), and 3) suggestions to the observer on how to correct an inappropriate response. Invalid data will be flagged for verification in the database. A raw message log will exist allowing administrators to see the original messages as sent by observers.

For all data entry, as a default, an identifying location ID, such as polling station, must be linked to observers who also have unique, numeric identifiers.

(Desired but not required) System could also explore options for future observer data collection through APIs

Data Export Options

The data management system should be able to export datasets in a common data format (guidance on the format to be provided upon request), including but not limited to:

1. (Desired but not required) Possibility to select individual observer records and export only those. Selection of the records would be made through an API.

Data Import Options

The data management system should be able to import the following datasets:

1. (Desired but not required) Possibility to select individual observer records and import only those. Selection of the records would be made through an API.

DATA REPORTING MODULE

Data Reporting: Who have we heard from? Who is missing and where?

For administrators, two 'views' are needed to assess who has reported, who has not reported, and the reporting response rates by geography:

1. Overview by reporting period:

The system should provide an overview of how many reports are complete, incomplete (i.e. a report with some question responses missing), and missing (i.e. no question responses have been submitted). The three types of data should be represented as three different colors and visualized in bar chart format. In addition, each reporting period should provide a visualization of the current overall response rate, enabling data managers to see how many reports are outstanding and where.

2. Response rates by the relevant geographic division

For each reporting period, there should be overall (i.e. national) numbers provided by status (complete, incomplete and missing) as well as numbers provided by relevant geographic sub-divisions. Response rates should be reported as both raw numbers and percentages.

The database should allow end users to apply filters and conduct searches for data. Database operators should have the ability to sort and filter polling station records by relevant geographic subdivisions, by message "status" (Complete, Incomplete, Problematic and Missing), and by sample (customized sub-groupings of observers set by the front-end user). These functions enable operators to rapidly identify the locations of observers who should be contacted regarding the status of their reports. Operators should also have the possibility to search by polling station code with additional searches possible based on observer/supervisor information such as phone number or unique ID.

Two-Way Communication with Observers

The database should allow for two-way communication with observers via SMS with bulk or blast sending features based on criteria. This includes not only the ability to automatically send confirmation of receipt messages as described under **DATA ENTRY**

MODULE, 2. Data Collection via SMS, but also to send customized alerts and updates to observers and to target these messages depending on reporting status, geographic location or supervisor. The system should include functionalities to make the messages personalized.

Data Export Options
The data management system should be able to export datasets in a common data format (guidance on the format to be provided upon request), including but not limited to:
1. (Required) Functionality to export the SMS log.

Data Import Options
> N/A

DATA QUALITY MODULE

Data Integrity or "Logical Checks": Do I trust the data?
The database should allow for the implementation of specific logical checks (to be provided on an individual use case basis) to check whether question responses are consistent with each other or with other reports received for an individual polling station. A logical check dashboard should enable end users to filter and/or sort the data by one or more combinations of logical checks and to export the results into a tabular CSV or XLS file (see **Data Export Options**). The database should allow NDI and the local partner to run tests to determine if, for example, turnout (sometimes called "signatures") is more than number of voters registered; or whether the number of registered voters pre-recorded in the system and number of registered-voters reported by observers is different. Checks can also be used to determine whether the response to certain questions (such as "How many polling station officials were present?") is always greater than or equal to the response to another question (such as "How many of the polling station officials present were women?"). The dashboard should provide end users with the ability to verify the integrity of data on a question by question or issue by issue basis. Access to the logical check dashboard should be controlled on the basis of user type with the option to expand or restrict access based on these types.

Data Export Options
The data management system should be able to export datasets in a common data format (guidance on the format to be provided upon request), including but not limited to:
1. (Required) Functionality to export quality assurance criteria, including the list of criteria as well as the records that have been flagged per criterion and combinations of criteria.

Data Import Options
The data management system should be able to import the following datasets:
1. (Desired but not required) Quality assurance criteria or lists of logical checks for data integrity.

DATA SUMMARY MODULE

Data Summarization of Process Data
The system should be able to provide the following automated summaries of submitted observer data at any time:
- Distribution of observer reports received by status (complete, incomplete and missing) by region. This should include absolute numbers and percentages for all reporting periods.
- (Desired but not required) It should be possible to sort the reports received by region on the basis of highest to lowest or lowest to highest regional response rate. A bar chart of all response rates could be provided including an average bar to show how each region compares to the national average.
- Summaries of responses for process-type questions: for each question the percentage (and absolute number) of reports that provided each of all possible answers to the question and of those who answered the percentage (and absolute) that reported each answer choice.
- Summaries of incident reports - display frequency across categories and regions and sample types.
- Ability to filter analysis reports by region, lower level of geography (e.g., county/district), polling station type if applicable (i.e. capital, urban, rural) and, if applicable, sub-samples.

Data Summarization of Turnout and Results Data and Calculations of the Corresponding Margins of Error
The system should provide:
- A summary of turnout projections both as absolute numbers and percentages
- Corresponding margins of error for turnout at confidence levels of 95% and 99%
- Ability to filter reports included in the turnout projections by lower level of geography (e.g county, district or state), polling station type (e.g. capital, urban or rural) and, if applicable, sub-samples.
- A summary of results projections by candidate/party both as absolute numbers and percentages
- Corresponding margins of error for results at confidence levels of 95% and 99%
- Graphs or charts should be considered to better visualize projections of turnout and results, including how both have changed over time with the submission of more reports.

Data Export Options
The data management system should be able to export datasets in a common data format (guidance on the format to be provided upon request), including but not limited to:
1. (Required) All qualitative and quantitative data associated with a polling station must be exported to a single data table which includes every polling station in the sample(s). The export option should be available (in CSV or Excel format) to those with sysadmin privileges. This will allow for in-depth data analysis, in addition to the internal system summarization functionalities outlined below. More

specifically, the export should be structured in a tabular format where rows contain the information about each form/polling station and the columns are comprised of the variables/questions (i.e., polling station ID, geographic information, observer information, and question/answers, reporting period, and log information on: time record created, timestamp when message received or timestamp when information last edited, which data entry operator last edited each record). An example export can be provided upon request.

2. (Required) Functionality to export Incidents in CSV or XLS format. (Desired but not required) The export should include links to multimedia like photos so that these can be easily re-imported into Ushahidi or Carto platforms.

3. (Desired but not required) An export of the overall findings and results for each form per reporting period. Results should be structured as follows:
   - Question, nulls, frequency/num, %
   - Option A, frequency/num, %
   - Option B, frequency/num, %

Data Import Options

The data management system should be able to import the following datasets:

1. (Desired but not required) An import of the overall findings and results for each form per reporting period. Results would be structured as follows:
   - Question, nulls, frequency/num, %
   - Option A, frequency/num, %
   - Option B, frequency/num, %

OTHER

Other Desired Database Attributes

- Credentialed login defines access to key reports and administrative features.
- Log of records information on: time record created, timestamp when message received, which data entry operator last edited each record.
- User friendly: The system should allow observers and administrators to use the application easily, through advanced search functionality, including filtering results by date, observers, etc.
- Mobile responsive interface: The system should display correctly on mobile phones and tablets.
- User management and download rights: The administrator should have comprehensive control over users of the site. Administrator should be able to grant access to differential features including data analysis reports and download rights.
- Track user reports: The system should feature a tracking system, recording all reports made by an observer.
- Consider exports or built-in systems to allow easy generation of graphs, charts and maps to visualize and provide more extensive analysis of data submitted.

Security of the Database

<u>System Access</u>
- The system will provide strong role-based access control lists
- User account creation and role or permission changes must be highly controlled
- The system will log user-input activities

<u>Website Hosting and Other Features</u>
- If the system is hosted, establish documented protocols for secure web hosting (e.g. use of a web application security scanner; manual audit for logical vulnerabilities; limitations to secure remote access; segregation of development, test and live environments).

Offline Access & Information Center Desired Requirements

For offline manual data entry and data management, a physical communications center should allow for, but is not limited to, the following hardware to be linked/used by database:
1. Local database server which syncs with internet server
2. Standby production server as a failover
3. 10-20 networked data-entry computers
4. Printers