CHAPTER THREE: Monitoring Electronic Techonologies Used in Voter Registration

INTRODUCTION

Voter registration is vital to democratic elections. In many countries prospective voters cannot cast ballots unless their names appear on the voter list at a specified polling station or are otherwise verified as being included in the registry of voters. A proper voter registration process is thus a prerequisite to citizens being able to exercise the right to vote and the right to be elected. Voter registries are developed in different ways, and increasingly they employ electronic technologies. This creates a need to review the ways that the public and the political contestants can gain confidence in voter registration efforts through transparency and monitoring of electronic technologies used in the process.

Observation groups and political contestants that are evaluating a voter registration process will soon realize that voter registration is administratively complex and technically sensitive.[25] For example, where election officials generally respect voter eligibility requirements and follow the law and regulations for registration of voters, a significant number of voters nonetheless could find themselves excluded from the voter registry — and thus disfranchised — because of poor execution of the registration process. There are numerous examples where the production of the voters list was problematic because of the poor use of information technology. There are also examples, such as in the 1994 Dominican

[25] This Guide concentrates on IT in the voter registration process. For a discussion of monitoring the broader administrative and other aspects of voter registration, please see generally, Richard L. Klein, Patrick Merloe, Building Confidence In the Voter Registration Process: An NDI Monitoring Guide for Political Parties and Civic Groups (NDI 2001), available at www.ndi.org.

Republic elections, where the final voter lists were printed and distributed to polling stations based on a fraudulent manipulation of the database. As with monitoring technologies used in other aspects of the electoral process, evaluation of the use of technologies in voter registration provides valuable information on the quality and integrity of the election.

It is important to note that evaluating the use of technology in the registration process can be cost and time effective. While monitoring the use of electronic technology in voter registration may require detailed knowledge of specific technologies, developing an understanding of basic principles is important for deciding on monitoring approaches. Even if observation groups and/or political contestants do not have a capacity to evaluate in detail a specific technology or range of technologies being considered for application in the voter registration process, they should have a firm basis for approaching the issues and for determining what kind of assistance they may need.

UNDERSTANDING VOTERS LIST DATABASES

If the voters lists are electronic and not paper records, they are contained in an electronic database. The lists can be kept in some decentralized form, for example by election district or municipality, or they can be centralized into one national voter registry. In order to understand how election authorities are managing registration of voters and how they operate voter records, it is necessary to grasp the basics of how databases work and some terminology related to databases and formats of the voter data.

"Voter's Record" is all of the information related to the individual voter.

"Primary Voters List Database Data" is information that is required to be in the voter lists by electoral legislation (for example, first name, last name, date of birth, etc.).

"Secondary Voters List Database Data" is information that is not required by the legal framework, but is useful in overall administration of the electoral process (for example, assigned polling station, flags, record tracking data, etc.).

"Format of the Voter Record" will define the kind of operations that are possible with the data. Following are simple examples to illustrate this.

In case A, the voter record is divided into three columns. If the electoral authorities want to separate voters according to a specific criterion such as provence, for example, it would not be simple to do so.

Case A

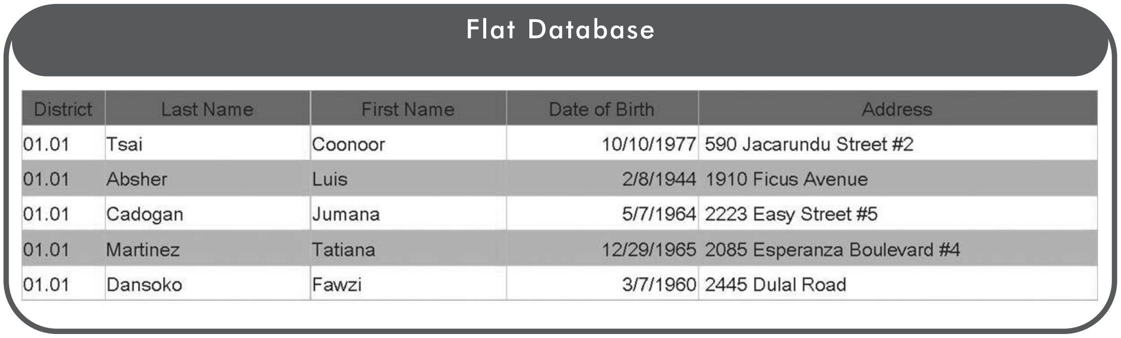
Name	*Address*	*Region*
Maria Chen	Main Avenue #13 Springfield, Sojob Provence	Eastern

In case B, it would be possible to separate voters based on several criteria.

Case B

First Name	*Last Name*	*Street*	*House*	*Town*	*Provence*	*Region*
Maria	Chen	Main Avenue	13	Springfield	Sojob	Eastern

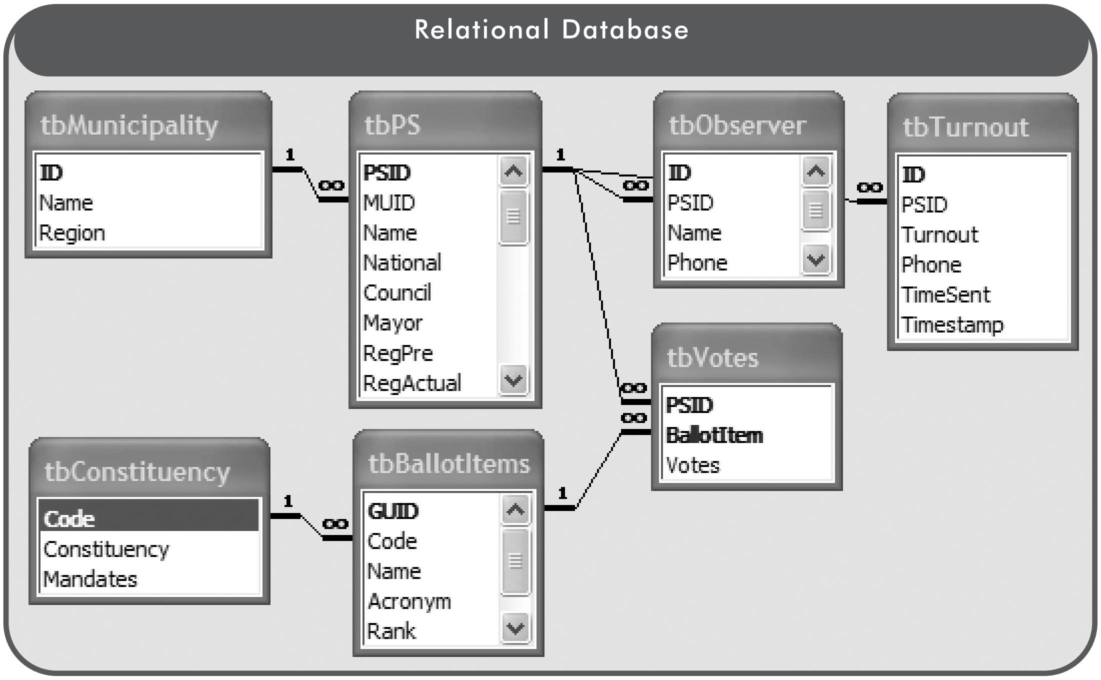
"Flat Databases" look like a spreadsheet. They have a simple design; each row represents one voter; columns contain information on each voter's first name, last name, date of birth, and extended address with complete geographical information. The data can be easily observed, but management and processing of the data is not practical. The redundant nature of some of the data increases the size of the data file, making it difficult to update and run queries.



Flat Database

District	Last Name	First Name	Date of Birth	Address
01.01	Tsai	Coonoor	10/10/1977	590 Jacarundu Street #2
01.01	Absher	Luis	2/8/1944	1910 Ficus Avenue
01.01	Cadogan	Jumana	5/7/1964	2223 Easy Street #5
01.01	Martinez	Tatiana	12/29/1965	2085 Esperanza Boulevard #4
01.01	Dansoko	Fawzi	3/7/1960	2445 Dulal Road

"Relational Databases" are designed in a more complicated way, in order to increase efficiency of the computing and data manipulation process. They have many tables that are related and "share" information. For example, it is very likely that the information about which polling station a voter is assigned will appear in a column that receives the information on polling station assignments from a different table.



"Database Product" is an output of the database containing a compilation of information available in a variety of formats intended for the end user. For example, a database product could be a printout of a final voters list or a webpage where a voter can correct his or her records or data for electronic poll books. To evaluate the product of the database, it is necessary to understand the architecture of the database, because the product does not indicate how the data were processed and whether there were technical flaws in compiling the list. For example, the exclusion of underage voters could have failed because the label in the database that marks underage voters was not part of the query that extracts the records of eligible voters.

"Database Exports" are electronic versions of some or all of the records in a database intended to be used by another database and thus not "intelligible" for people. The export can be described as an intermediary product.

"Database Design Requirements" are set by the election authorities and inform the specifications that are used by programmers to build the database. Requirements should be derived from the needs of the electoral process. It is impossible to build an adequate database without first understanding what kind of data are collected and used. Once the input of the data into the database initiates, changes in the database architectures are limited and risky. Adding or removing capabilities from the database is best done at the requirements phase of the process. A poorly conceived and poorly built database leads to repetition of input of records, an inability to properly manipulate records and corrupts the transparency of the database.

"Database Accountability" refers to the requirement outlined in the database design that, in addition to voter records, requires the database to keep records of changes, deletions and insertions for review purposes.

USE OF EXISTING RECORDS— TRANSFER OF RECORDS

When existing databases (such as civil registries) are used as a basis for creation of the voter list,[26] election monitors and political contestants usually do not have complete access to the "original" databases.Their access is limited to the voter list database. However, in order to understand the transfer process of the voter records from the original database to the voter list database, political contestants and election monitors should understand the following features of the original database:

- process of the data collection;
- management and update of the records;
- compatibility of the database with requirements of the voter list; and
- capability of the database to export the data and the features of the export.

[26] In some countries the civil registry is in fact a register of voters and does not involve creation of the separate voters list that is managed and updated by election officials—for example, Denmark and Sweden follow this model.

It is often impractical to use existing original databases as the voter list database. They are not built to serve as voter list databases, and they contain information that is not related to the voter list. Therefore, the transfer of records will not be a simple process of copying the original database. Data have to be prepared for the "receiving" (voter list) database; and they need to be exported in the format that the voter list database can receive.

In countries with a long history of voting and use of existing civil registries for the creation of the voter list, civil registry database design and integrated data management tools are sometimes utilized for efficient export of the civil registry records to the voter list database.

Even if the civil registry is well maintained and contains all of the data required for the voter list (including primary and secondary voter list data), the transfer of records to the voter list database can be troublesome and even create a fatal flaw in the process.

In countries where the use of existing records for compiling the voter list is a first time occurrence, it is common for there to be numerous problems with the process. These problems multiply in cases of corrupted records, inadequate maintenance of the data, interrupted management of the dataset and in translation/transliteration of records in different scripts and languages.

Data Migration Process:

Where the civil registry is used as the basis of the voter registry, moving information from the civil registry into the voter registry will involve a data migration process. Data migration between these two systems, built in different ways to serve different purposes, can present a number of challenges. Differences in data contained within these two systems present the first challenge. Monitors should ask whether the information in the civil registry is sufficient to cover the primary and secondary data required for elections. Data migration can also be compromised by technical differences between these systems, such as differences in database design, software used and field formats. Migration must be careful to avoid losing relationships, primary/foreign keys and character sets.

Formatting of Fields and Records:

Every database has a defined format for each field. This information is integrated into the database. Fields that contain geographical locations will have different parameters than fields that contain dates or "flags." If the format of the field is not properly transferred into the data export, the receiving database will have difficulty recognizing these fields and may interpret them incorrectly.

The format of the records (e.g., how the individual records are divided in columns) will dictate what operations are possible with the records. For example, if address fields are not properly structured, it will be impossible to automatically assign the voter to a specific district or polling station. The process would have to rely on manual checking of the records or involve some type of software that would recognize addresses and assign the proper location code. In the case of automated recognition of addresses, error rates may be significant and correction efforts must be planned.

COUNTRY NOTE:
Ukraine 2007 - Incompatibilities of Databases Required 11 Million Manual Re-Entries on Voter Lists

Ukraine held early parliamentary elections in 2007 as a result of a protracted political crisis. Amendments to the election law required the voter registry for the 2006 elections be sent by the Central Election Commission (CEC) to 679 Working Groups around the country for them to merge the 2006 voter lists with databases from 10 state agencies and otherwise update the voter lists. Incompatibilities between database software resulted in the manual re-entry of information for approximately 11 million prospective voters. The Working Groups delivered draft voter lists to the District Election Commissions, as required by law, without passing them back to the CEC for it to create a national voter registry and/or to conduct verifications as was done in the 2006 elections. There was a short period for the public to scrutinize the voter lists and file for corrections, and the corrections process was not well publicized. While the quality of the voter lists varied around the country, double and multiple entries occurred in significant numbers on the 2007 lists, while other problems led to exclusions of qualified voters from the lists, thus creating opportunities for illegal voting as well as disenfranchisement. These factors led to lower public confidence in the voter lists and a general assessment that the 2007 voter lists were not as accurate as those of the previous year.

Sources: "Preliminary Statement of the NDI International Observer Delegation to Ukraine's September 30, 2007 Parliamentary Elections", NDI (1 October 2007); "Statement of Preliminary Findings and Conclusions on the 30 September 2007 Parliamentary Elections in Ukraine", OSCE, et al. (1 October 2007); "Pre-Election National Monitoring Report", OPORA (Support)(27 September 2007).

Unique Identifiers:

Unique identifiers are also called "primary keys." These are entries in the databases that serve to unmistakably identify a specific set of information, for example, a voter. Rather than linking different pieces of information to the voter's name, a primary key is assigned to the voter so that the database effectively identifies individual voters. These keys must have a standardized, distinct and well defined format so that the database can properly maintain relationships between different pieces of information.

Software Compatibility:

Software of the original and receiving databases have to be compatible so that the export of the data from the original software can be imported into the receiving database without loss of individual pieces of information or relationships between the data. One of the most common problems is different language schemes and definitions of character sets between exporting and receiving databases. Databases might operate with different systems that define letters and numbers. Even within the same language script there could be differences in use of coding standards. The situation becomes more complex if the script of the exporting database has to be transliterated into a different language script for the receiving database.

COLLECTION OF DATA

Creation of a voter list that is a "voter registry" independent from other registries (such as, the civil registry) involves collection of voter data by election authorities. However, rarely is an independent registry truly independent. There are almost always aspects that depend on the work of other institutions (e.g., the Ministry of Interior that issues ID cards or other proof of citizenship or the Transportation Department that issues driver's licenses, which are used by voters to prove their eligibility). Also, it is not unusual in these circumstances for the creation of the independent voter registry to be a one-time occurrence, and updates to be processed by some automated mechanism that requires sharing of data with institutions that are issuing birth, marriage or death certificates or some other means of recording the status of citizens.

It is important that monitors understand all manners of populating the voter database and recognize there will inevitably be some degree of error in creating the voter list. Database design and management processes should include "built in" tools to tackle this issue, but monitors should also look into what steps are taken to minimize, uncover and correct error.

This section will discuss issues related to the monitoring technologies used in the creation of the voter list, irrespective of whether the creation will be a one-time occurrence or continuous or periodic exercise, or whether it will be a voter-initiated or state-initiated process. What they all have in common is that the voters' data are not immediately recorded as electronic records in a central voter registration database and that fairly complex and sensitive operations must be used to collect and process these data.

Whether the collection of the data is done by direct or indirect recording, it is important to determine what type of information is being captured and whether this reflects the requirements of the legal framework. If election authorities are collecting data beyond what is required by the legal framework, this must be properly justified or discontinued. If election authorities are collecting data that will be shared with other governmental institutions, this should be disclosed.

Direct Recording:

Direct recording involves creating an electronic voter record at the moment and location when the voter (or his or her proxy) submits the data to the election officials in accordance with the law and regulations. In direct recording, voters do not fill out a form that will later be entered into the voter database by scanning or data entry in some remote location. Rather, their data is captured directly at the registration point using electronic equipment.

Development of the System. Observation of the direct recording technology must start at the point when election officials are developing specifications for hardware and software requirements. These requirements must match the model of the registration exercise — for example, mobile versus stationary registration points or a large number of points versus centralized locations. Equipment requirements will differ if the equipment has to be transported or if it is stationary, if it relies on infrastructure (such as, electricity or networks) or if it is designed to work without infrastructure (for example, to run on batteries).

Software. Electronic records that the registration equipment creates must be compatible with the voter registry database so that records can be easily and accurately transferred to the central database. Principles discussed above, under "transfer of existing records" apply here too.

Testing. Direct recording equipment should be properly tested before it is deployed. Tests should be performed following the "end-to-end" principle, meaning that the complete process is simulated with actual components of the system and exact copies of the software in an environment that is similar, if not exactly the same as, the type where the equipment will be utilized. A complete testing and monitoring process requires recording data of people involved in the test at the actual registration points and transferring this data to the central database. In addition, "load tests" should be performed to gain a better understanding of how the equipment deals with the expected number of transactions and whether projections of the number of processed voters are realistic. Tests should also be conducted concerning how the database responds to malfunctions and problems.

Tests are performed not only to verify functionality of the equipment and the process, but also to examine usability of the system, both from the voters' and election officials' point of view. Beyond the functioning of equipment, authorities should solicit the opinions of all those involved in testing - simulated voters, officials handling equipment, supervisors and others. Monitors from political contestants and election observation groups should be allowed to provide input regarding any concerns they may have before tests are designed, review and ask questions about the testing procedures before they are conducted, witness all testing and be provided timely access to the opinions of all actors involved in the testing.

It is not expected that monitors from election observation groups or political contestants will perform these tests; however, they need to be able to evaluate how the testing was performed. Testing of the systems is part of the electoral process. It requires that election officials have a clear test plan and that testing and outcomes are recorded and shared with monitors in a timely and understandable manner.

If tests are performed on a smaller scale, for example on a small sample of equipment, the tests are considered design tests or model tests. Performance tests are those that test the complete set of equipment. If the election officials do not perform a full scale performance test, it is necessary to establish criteria by which a sample of the equipment will be tested. The sample should be on a proper statistical probabilistic sample, where every piece of equipment that will be deployed to registration points has the same chance to be selected. Tests should not include just "the first 100 pieces of equipment delivered" or other arbitrary criteria because such tests have proven to be unreliable indicators of how the full set of equipment will perform.

Monitors from the political contestants and observer groups should be allowed to review sampling methodology. Monitors from observation groups and political contestants must thoroughly understand the system in order to evaluate whether performance tests can be reduced to test a sample of the equipment. Sometimes it is absolutely necessary to conduct full scale tests, especially if the equipment requires calibration and fine tuning (such as bio identification systems like fingerprint scans) or if it is impossible to troubleshoot problems once the equipment is deployed.

Accountability. As with every other aspect of the electoral process, direct recording of voter data should follow the principle of accountability. This means that every sensitive action should be recorded and stored to provide opportunities for possible examination. Since electronic records are not accessible to the public, individual voters cannot verify whether the equipment recorded their data properly. Therefore, direct recording registration systems must provide each voter with proof of her or his submission of their data. This proof can be a printout of the voter's record or some other type of receipt or certificate. Voters thereby are given an ability to prove their involvement in the registration process, which is usually needed in order to seek remedies should they discover errors or omission of their data.

In addition to the receipt that confirms submission of the data, the voter should receive a unique number for the transaction that will serve as an identifier. The receipt and identifier can aid voters in exercising their right to check the preliminary voter list and demand corrections if data is erroneously recorded or if the voter is somehow omitted from the list. The receipt and identifier also can aid election observation groups and political contestants to conduct independent verification exercises with the consent of registered voters, who agree to participate in such efforts.

Security, Back Up and Data Transfer Procedures. Security procedures should address two principal issues: (1) security of the data regarding unauthorized access and manipulation of data; and (2) security regarding potential loss and corruption of data. Election authorities should have defined security procedures that are made available for review by monitors from observation groups and political contestants. Monitors would not obtain security codes granting them access but would be able to comment on whether the procedures themselves seem adequate.

To ensure adequate security, data must be protected with technical and organizational solutions, and election officials should employ both methods to secure the data. Technical solutions are built in to the equipment and limit access to authorized election officials. Equipment must be tamper resistant or at least tamper evident. Technical security solutions should also have clearly identified access levels - not all of the officials should have access to all of the data and processes. Organizational solutions are a set of rules that election officials must respect to protect access to the system.

In order to protect data captured at the registration points, election officials must design a reliable back up process. Back ups have to be regular, scheduled and documented. Also, backed up data should be stored independently from the direct recording equipment, so that in case of malfunction of the equipment and loss of the original data, back ups are preserved. Storage and management of the back ups should also be included in design security procedures.

Monitors from political contestants and observer groups should also be allowed to evaluate procedures for the data transfer. Data transfer can be physical (e.g., by moving memory cards from the direct data capture equipment to the central database) or through a computer network. Data transfers are sensitive points in the process since they pose a challenge to protection of the data by introducing elements of uncontrolled environments. Monitors should be allowed to accompany physical transfers or evaluate such transfers based on sampling techniques and should be allowed to evaluate transfer of data by networks through reliable techniques, such as comparing data sent from a particular machine or registration center (or sample of machines or centers) to corresponding data recorded centrally.

Development, Delivery, Maintenance, Troubleshooting and Service of Technologies. Ensuring the proper functioning of the direct recording equipment and related technologies—like every other aspect of election administration—is the legal responsibility of the election authorities. In effect, the election authorities have a duty to properly discharge the obligation of government to provide genuine democratic elections to the citizens, including to the voters and to those standing for election. It is common that election authorities outsource development and production of the technologies to independent companies, and they often rely on the private companies (that many times are foreign entities) to deliver, maintain, service or otherwise troubleshoot problems with the technologies. This normally creates a legal contractual relationship between the election authorities and equipment producers (vendors) and/or servicers. However, that legal relationship is subordinate to the election authorities' legal obligation to citizens, which is set by the country's constitution, electoral law and often reinforced by international human rights obligations.

The role of the equipment producers and/or servicers and the capacity of the election officials to service equipment is an important consideration in ensuring electoral integrity. The importance of building capacities of election authorities and avoiding over-reliance on vendors is essential to meeting a government's obligations to organize genuine democratic elections. Delivery of equipment should be complemented by the transfer of know-how to electoral authorities to effectively service the technologies, or electoral authorities must ensure that producers and/or servicers are in-country and in position to provide effective service that allows the technologies to perform according to the registration plans. Otherwise, the entire voter registration process can be jeopardized.

Contracts therefore should be open to scrutiny by observation groups and political contestants.

COUNTRY NOTE:

Nigerian Elections 2007 - Use of Electronic Technologies in Voter Registration

While the Nigerian electoral act prohibits electronic voting, the Independent National Election Commission (INEC) decided to employ direct data capture (DDC) devices to create an entirely new voter registry for the series of elections held in 2007. DDC technology would have enabled officials to electronically enter and store information about each voter who appeared at registration locations and then transfer the information to a computer database. Election authorities would then have been able to conduct various checks to ensure the integrity of voter lists, for example, to identify duplicate records and thus prevent double voting. However, the INEC's very tight and optimistic timetable proved not to be realistic. INEC expected to procure from three companies a total of 33,000 DDC machines by early November in order to complete registration of an estimated 70 million eligible voters by the December 14 legal deadline. At the beginning of registration only about 1,000 DDC machines were operational, and due to a number of factors, including delayed payments to the vendors, the 33,000 machines were not in place until mid-January. Only about 5,000 of the machines were voter registration devices, while the majority of machines used were laptop computers with digital cameras. In addition, registration staff apparently did not receive sufficient training on the use of the DDC devices. The batteries provided had a short life span and recharging facilities were limited in number, often rendering the DDC devices unusable. The printers frequently jammed, and there were shortages of ink. A manual registration process had to be used as back-up. The result was significant delays beyond legal deadlines, a problematic correction period, which led to likely disenfranchisement, and opportunities for illegal voting due to inaccurate voters lists. While aggregate registration figures were made public, there were questions about the large volume of registrations in the final phase of the exercise. Public confidence was further compromised because significant access to the voters list was not provided to political parties or domestic and international observers prior to election day. Eighteen political parties joined in a court challenge concerning noncompliance with legal provisions on voter registration.

Sources: "NDI Final Report on Nigeria's 2007 Elections,"; "Nigeria Final Report: Gubernatorial and State House Elections 14 April 2007 and Presidential and National Assembly Elections 21 April 2007," European Union Election Observation Mission.

Obligations of the producers and/or servicers after delivery of the products should be clearly defined by contracts that carry an appropriate level of guarantee that the producer will indeed effectively service the equipment. The contracts should address obligations to effectively remedy breakdowns of equipment due to design flaws, as well as due to operation in high temperatures, high humidity, exposure to sand particles, failures of batteries needed to operate equipment as specified; and the ability to rapidly provide replacement parts and otherwise ensure equipment performance. The schedule for delivery of equipment needed to meet the election authorities' voter registration plan should be verified against the producer's available inventory and production schedule (including obligations to deliver equipment and technologies to other countries). All of these issues have had serious negative effects on voter registration processes and must be taken into account.

It must be expected that something will go wrong during registration of voters. Tests should help to identify and minimize weak points and reduce malfunctions, but officials must expect and plan for problems. A bigger problem than failure of some components is not having an effective response plan. Response plans must be clear and documented. They must define response steps, response times and roles. If the response involves the equipment producer or another contracted company, this should be clearly defined in valid contracts. Such response plans should be made available to observer groups and political contestants, with opportunity for their comment. This is an important point for genuine transparency and confidence building.

Training. Election officials who perform voter registration should be trained in verification of the voter's eligibility, in how to properly record the data and in how to otherwise operate the equipment. They must understand the functioning of equipment (technologies) on at least a basic technical level in order to identify problems, to be prepared to correct them on the spot, if possible, and to request appropriate assistance and service.

The training should be in line with standard training requirements - trainings should be thorough, mandatory, standardized and include simulations of normal procedures and responses to malfunctions. Monitors from observer groups and political contestants should as a best practice be allowed to review training plans and materials before they are employed and to provide comments. Monitors should in any case be allowed to attend and observe training sessions to build confidence in how officials will be prepared to use technologies during the voter registration process.

Indirect Recording:

Indirect recording of voter registration data employs collection of data through non-electronic means, which is later processed and electronically recorded into the voter list database. Data are first collected on paper forms and then entered into computer systems either by manual data entry or scanning.[27] Scanning technologies as well as manual data entry present a number of challenges to electoral integrity. Monitors from observer groups and political contestants should be allowed to witness end-to-end testing or performance testing of scanning technologies, as with direct date capture technologies. Issues related to development, production, delivery, servicing, maintenance, troubleshooting and training discussed above also apply to indirect recording technologies.

Forms and Data Sources. There are two principal categories of data sources for indirect recording of voter data. The most usual source is forms created for the purpose of data collection. However, there are cases where election authorities use existing paper records, such as index cards or previous non-electronic voter lists. The primary difference is whether the election authorities are creating a new data collection from scratch or relying on existing paper records. If starting from scratch, the election authorities can (and should) design their data collection process and forms with the database and their information needs in mind. If they rely on existing paper records, the election authorities will have to be more creative in how they digitize the existing information and introduce it into the database. These processes are vulnerable to error in different ways.

Forms for capturing voter information must be designed to be compatible with the format of database records. Improper design of the fields on the forms, for example, leads to problems (or at least complications) when merging the data recorded on the forms into fields in the database. Form fields must be properly coded to speed up and facilitate data entry. It is also advisable to code the forms with a unique number in order to create a paper audit trail.

In terms of layout, forms have to take into consideration the applicable data entry method — a form that is prepared for scanning is different than one that will be used for manual data entry. The scanned forms have to be machine readable, while the manual data entry forms have to be human reader friendly. In either case, the forms must be understood by the person filling them out — whether

[27] In exceptional cases, data can be gathered with some other type of electronic record that would still need additional processing. An example would be typing the data into a word processing program that is not compatible with the voter list database and then "re-recording" the information into the database. In such processes there are risks that data could be corrupted, while the original record could be easily lost.

that is an election official or a prospective voter. A form that is easily readable by a scanner or data entry person that nonetheless is likely to lead to improper or incomplete information presents a major problem for the integrity of the registration process.

Forms therefore also should be available for review and comment by monitors from observer groups and political contestants. Having confidence in form design will provide a basis for building confidence in the training of the election officials who will complete the forms and/or in voter education — both of which are additional elements of the registration process that should be open to monitors.

Data sources (such as prior voters lists or index cards) that are not designed for data entry will likely present problems for scanning. If such sources are to be scanned, proper testing should be performed to determine a practicable entry method. If they are to be entered manually, it is advisable that the information on the paper be marked with field codes in a pre-entry process, especially if the layout of the forms is not data entry friendly.

Both types of data sources might require reformatting, converting and coding of certain kinds of information, for example conversion of dates from different types of calendars or coding of geographic areas.

Understanding the format of the data sources is useful for anticipating the kind of challenges that the source is likely to pose for the data entry process. Knowing how and why the data was prepared and reformatted to accommodate the data entry will also help. There have been cases when data entry has failed because of poor preparation of the data source. Therefore, plans for such data entry should be open to review and comment by monitors from observer groups and political parties.

Manual Data Entry. Entering voter data into the voter database is a large undertaking because election authorities have to enter millions of records. The capacity of the data entry system is therefore an important issue. Planning of the data entry system should involve testing such capacities. Testing would involve load tests (to determine how much of the data can be processed in a given time period), performance tests (to determine if the data entry interfaces respond well, if the networks are stable and the server can deal with large numbers of entries) and functionality tests (to determine if the interface design is appropriate and does not contribute to data entry errors). These tests, including review of testing results and recommendations, should be open to monitors from observer groups and political contestants.

Every data entry system should have different levels of access for operators, supervisors and administrators. Operators should not be able to access any records except those that they are currently entering. Supervisors and administrators should have higher levels of access, and their involvement should be necessary to correct and edit the data.

Every data entry system should include post-entry checking. This means that printed listings of the data should be given to a group of editors (verifiers), who would compare entered records with the data source (e.g., forms). Any errors should be marked and their findings passed onto supervisors and administrators, who take corrective action. This measure reduces greatly typing mistakes and other human errors. Another way to insure the quality of data entered is double entry. Double entry involves entering the same data by two separate groups of operators in two separate operations. Data are then compared and records that don't match are marked for inspection. Reports of the percentage of mistakes identified and corrected should be available to monitors from observer groups and political contestants.

As in every database operation, the complete audit trail should be recorded in the data entry software. Recorded information should include the time of the creation of the record, its source, the operator, each change and who authorized each change. Monitoring of such information could be done by experts or independent audit firms contracted by observer groups or political contestants and charged to evaluate whether proper procedures were followed in production of final voter lists.

Scanning — Optical Mark Recognition and Optical Character Recognition. The advantages of employing scanning technology over manual recording of voter data for voter registration are obvious — they significantly reduce the need for large infrastructure and data operators. However, scanning machine error rates have to be known in advance, and plans are needed to identify and correct errors. Human "reading" of scanned data, back up manual entry and correction of the records must be considered.

VOTER REGISTRATION DATA ENTRY IN BOSNIA-HERZEGOVINA
SCANNER VS. MANUAL ENTRY

	Scanner	Manual Entry
Registrations	3,500,000	3,500,000
Forms per hour	4,500	60
Work hours per day	16	12
Forms per day per person/scanner	72,000	720
Scanner/person days required	49	4,861
Number of scanners/persons	5	100
Total forms per day	360,000	36,000
Days to complete data entry	10	96
Error rate	0.10%	2.00%
Forms to be re-entered	3,500	70,000
Re-entry days	>1	2

Source: Final Report, OSEC Elections Assessment Team, Mission to Bosnia and Herzegovina, January 30 1996

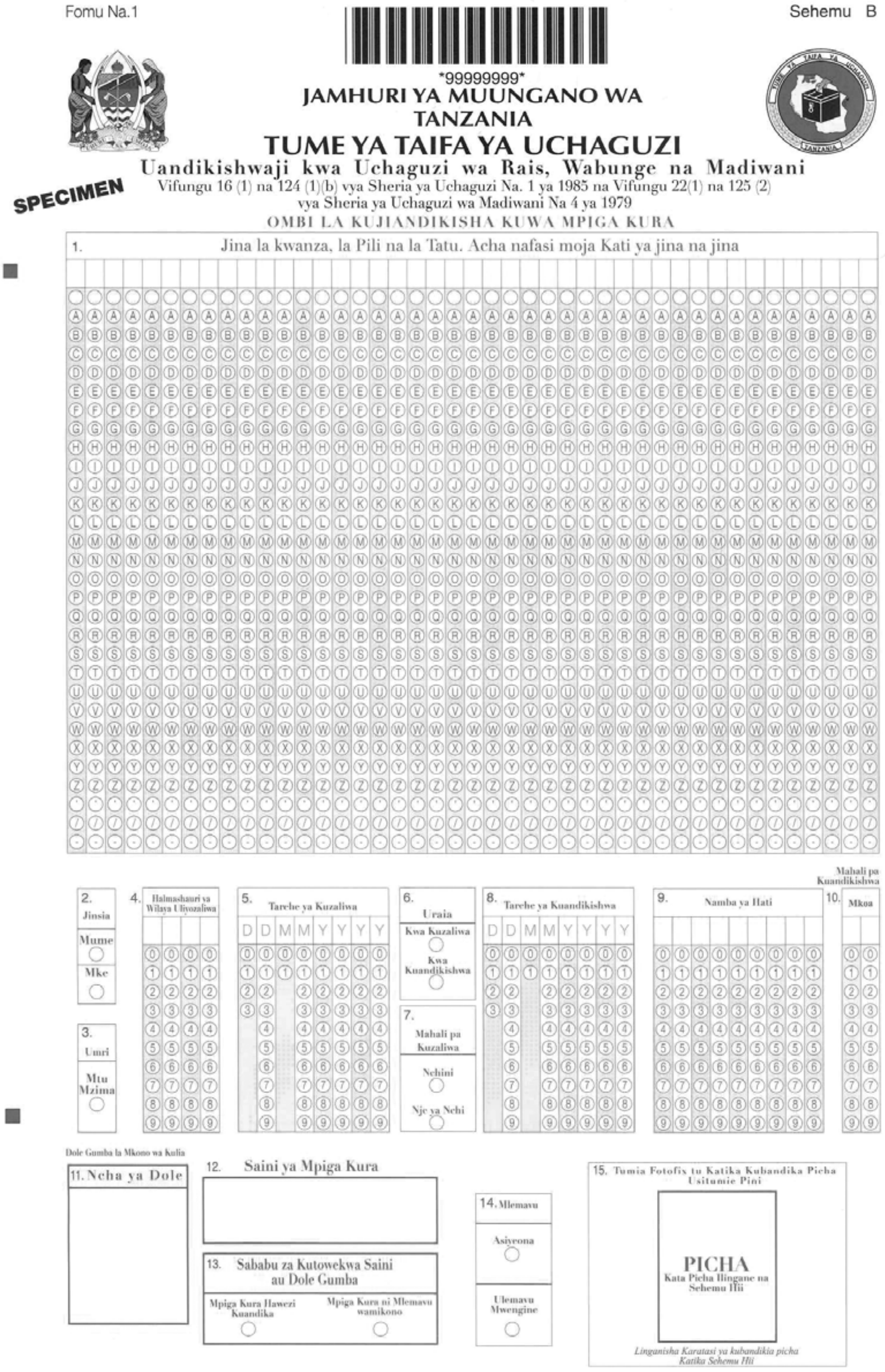
More than manual data entry, the quality of the scanning will depend greatly on the format of the data source. A data source that was not formatted for scanning will likely create so many corrupt records that the whole exercise could well be futile. The format of forms prepared for scanning is not user friendly for human readers because it is designed for the scanner and scanning software.

OMR systems are more accurate than OCR systems.[28] The OMRs recognize marks entered on forms, while OCRs are used for processing written data. To improve accuracy of the scanning process of OCR applications, it is advisable to numerically code as much information as possible; limiting input to just digits reduces the number of character options and therefore opportunities for misinterpretation.

It is possible for election authorities to use scanners without OCR software and create images of the form - these systems are much cheaper than those equipped with OCR software. Scanned images

[28] Please see Chapter 2, "Optical Mark and Optical Character Recognition," for further discussion of this subject.

Optical Mark Recognition (OMR) Form
Source: National Election Commission of Tanzania



Fomu Na.1

Sehemu B

99999999

JAMHURI YA MUUNGANO WA TANZANIA

TUME YA TAIFA YA UCHAGUZI

Uandikishwaji kwa Uchaguzi wa Rais, Wabunge na Madiwani

Vifungu 16 (1) na 124 (1)(b) vya Sheria ya Uchaguzi Na. 1 ya 1985 na Vifungu 22(1) na 125 (2)
vya Sheria ya Uchaguzi wa Madiwani Na 4 ya 1979

SPECIMEN

OMBI LA KUJIANDIKISHA KUWA MPIGA KURA

1. Jina la kwanza, la Pili na la Tatu. Acha nafasi moja Kati ya jina na jina

○ Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ Ⓕ Ⓖ Ⓗ Ⓘ Ⓙ Ⓚ Ⓛ Ⓜ Ⓝ Ⓞ Ⓟ Ⓠ Ⓡ Ⓢ Ⓣ Ⓤ Ⓥ Ⓦ Ⓧ Ⓨ Ⓩ ' / .

Mahali pa Kuandikishwa

2. Jinsia

Mume ○

Mke ○

3. Umri

Mtu Mzima ○

4. Halmashauri ya Wilaya Uliyozaliwa

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

5. Tarehe ya Kuzaliwa

D D M M Y Y Y Y

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

6. Uraia

Kwa Kuzaliwa ○

Kwa Kuandikishwa ○

7. Mahali pa Kuzaliwa

Nchini ○

Nje ya Nchi ○

8. Tarehe ya Kuandikishwa

D D M M Y Y Y Y

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

9. Namba ya Hati

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

10. Mkoa

⓪ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨

Dole Gumba la Mkono wa Kulia

11. Ncha ya Dole

12. Saini ya Mpiga Kura

13. Sababu za Kutowekwa Saini au Dole Gumba

Mpiga Kura Hawezi Kuandika ○

Mpiga Kura ni Mlemavu wamikono ○

14. Mlemavu

Asiyeona ○

Ulemavu Mwengine ○

15. Tumia Fotofix tu Katika Kubandika Picha Usitumie Pini

PICHA
Kata Picha Ilingane na Sehemu Hii

Linganisha Karatasi ya kubandikia picha Katika Sehemu Hii are then transferred to a central location and processed by higher quality computers using OCR software, which can produce records with fewer errors. An audit trail of the scanning data entry is provided by the image of the registration form or other paper data source.

No matter what kind of database is used to store images of the form, it must connect the image and the paper record to provide accountability. If scanners without OCRs are used, a system for manual marking of the forms should be developed. Such a marking process would assign a unique identifier to the paper form that is recorded with the form's image (usually within the image file). If OCR or OMR is used, those filling out the form should be given sufficient instruction on how to complete the form in a way that will minimize error.

Forms that were not clearly and completely processed must be re-checked and entered manually. For that purpose, OCR software should have a built in error detection function and should be able to separate corrupt scans. Even forms that are properly scanned may require significant manual review to verify character interpretations. Most OCR software includes verification tools that allow a human operator to quickly view each character the computer wasn't able to match perfectly and compare it to the original scanned image.

VOTER DATABASE REQUIREMENTS FOR AUDITABILITY

Evaluation of the voter database should aim to review two interconnected aspects of the functional database: (1) database design; and (2) database management. It is not possible to separate these two elements because the database must be designed to address management requirements and some management policies are designed to address database structure. Election authorities should build evaluation and testing into the voter registration plan, and monitors from observation groups and political contestants should be allowed to review and comment upon the plan for such evaluations. Monitors should also be allowed to witness the testing and evaluations, or at a minimum be allowed to review reports presenting results of testing and evaluation. In addition, as discussed below, monitoring by observation groups and political contestants

COUNTRY NOTE:

Indonesia 2004 - Voter Registration Using Optical Character Recognition (OCR) Scanners and NGO Voter Registration Audit

In 2004, Indonesia held its first presidential elections, and second legislative elections, in its democratic transition process. A voter registration exercise was conducted across the country in April and May 2003 in preparation for the elections. The exercise faced the challenge of reaching the country's more than 17 thousand islands and over 150 million voters. Voter registration officers visited households during this period, capturing data on all eligible and ineligible citizens on optical character recognition (OCR) forms. The forms were processed at 45 state statistical bureau offices in all 30 provinces. A total of 92 scanners were used running 23 hours a day, seven days a week. During limited trials the scanners were 93 percent accurate at letter recognition and 97 percent accurate at number recognition. In February 2004, JURDIL Pemilu 2004 (The University and NGO Network to Monitor the 2004 Elections) and one of its member organizations LP3ES (Institute for Social and Economic Research, Education and Information) sent out 400 observers to conduct an audit of the voter registry. The audit used a statistical sample, comprising 5,592 voters from 375 randomly selected villages in 12 of the country's 32 provinces. It found that the registry contained approximately 91 percent of eligible voters, with some variance among provinces (81% - 96%) and a difference between certain marginalized groups (minorities, displaced persons and those in conflict or very isolated areas) versus the general population (86% v. 92%). The audit found a small incidence of persons on the list who did not exist; however, it identified a significant number of errors in dates of birth, which it noted could have resulted form many people not knowing their exact birth date. In part due to the audit, a follow-up voter registration exercise was conducted that increased the number registered voters to over 95 percent (an increase of several million voters).

Source: "Consolidating Democracy: Report on the UNDP technical assistance program for the 2004 Indonesian elections," United Nations Development Program (New York, undated); "Voter Registration Audit Report," JURDIL Pemilu 2004 (10 March 2004).

should extend to being allowed to examine policies and procedures concerning security of the technologies and the voter list itself, and to conduct independent audits of the voter list.

Evaluation should start with a review of the functional requirements that the election authority provided to programmers of the database. If the election authority does not define the functional requirements, it is likely that the programmers will create a database that is efficient in terms of computation and manipulation of data, but probably does not accommodate peculiarities of the electoral process. This potential shortcoming in the election authority's planning would create a circumstance where the technology will impose requirements on voter registration and force the electoral process to accommodate the information technology (IT), rather than vice versa. Definition of the functional requirements is best done upon discussion with observer groups and political contestants, including review of "peculiarities" in the country's legal framework. Such input can provide important insights, and the participation can build confidence in the process.

In principle, the voter list database should be designed to meet the following requirements:

Primary Voters List Database Data - The proper basis for voter registration and voter data management is the legal framework for the election process. The legal framework will determine the types of data that need to be included in the voters list. These data may go beyond names, date of birth and addresses, if the legal framework requires information beyond such basic voter data. Additional voter information could be required for voters who vote abroad or in the military service or who vote by absentee ballot, or for voters who are excluded because of rulings of mental incompetence or penal reasons. The database must accommodate these provisions.

Secondary Voters List Database Data - The voter database rarely includes only the basic voter information required by the legal framework. In order to administer elections, election authorities need to integrate more data into the database to ensure proper management of voters. These data include information such as assigned polling stations, different coding information and record tracking data. The content and types of secondary data depend on the management policies of the electoral authorities and consequent requirements.

Accountability - Records or information should never be deleted in the database. Instead of deleting records, databases should be designed to have "flags" that will mark that the record as "deleted" or changed. Following the same principle of accountability, changes in the database have to be recorded, with information about who changed the data and who authorized the change. This is called the "*Audit Trail*" - that is, the record of changes in the voter list database.

The "*Audit Trail*" is important for resolving efficiently and accurately disputes that may be raised by prospective voters in the claims and objections period. The database should accommodate timely dispute resolution processes.

Security and Access - Security evaluation of the database should identify sensitive points in the process of adding, updating or deleting records as well as overall safety of the records. This includes the physical security of the premises where the database is housed. In order to address security concerns, election authorities must establish technical solutions and organizational policies that will prevent unauthorized and undetected manipulation of the data. Database design must have defined access levels that are reflected in the database. Responsibilities of operators, supervisors and administrators must be defined and transparent.

Monitors from political contestants and election observation groups should be allowed to review policies regarding the overall safety of the records and should be allowed to review procedures that election authorities established for safe storage of data, back ups, transfers and other related matters. This can be done without compromising the security of the database, and such reviews add significantly to confidence in the voter registration process.

Compatibility - In cases where the voter database is developed by using preexisting records, or it is developed to deliver data in an electronic format to another database (for example electronic poll books), the design has to define carefully how the database will effectively interface with the databases with which it must interact.[29]

Overall Database Structure - Beyond specific requirements of the database for voter list purposes, evaluation of the database should assess the database structure. This includes review of relations between different data and tables, coding and categorization of the data, application of primary keys, definition of fields, and format of tables, records and fields.

Content Testing - Conducting tests of the content of the voter list is a step beyond monitoring the design and functioning of the information technology used in creating the list. These tests examine the electronic voter list (or often a copy of it) to identify errors, such as duplicate records, records with missing data, records that show ineligible persons were entered onto the list or voters assigned to the wrong constituencies. Computer tests can also identify trends in the voter list data that may raise questions about the representativeness

[29] For more on compatibility issues please see "Use of Existing Records - Transfer of Records," above in this Chapter.

of the list, which could indicate that certain population groups are over or under represented (e.g., gender, age, language or ethnic groups or people from certain geographic regions). This could be the result of manipulation of the database, errors in data entry, manipulation in data collection or faults in the registration process. All of these possibilities call for remedies, from removing duplicate records to extending the claims and objections period for list correction to even reopening the registration process.[30]

[30] "Computer Tests" of the voter lists are described in Richard L. Klein, Patrick Merloe, *Building Confidence in the Voter Registration Process: An NDI Monitoring Guide for Political Parties and Civic Groups*, 32-34 (NDI 2001).