A WORKBOOK BY THE NATIONAL DEMOCRATIC INSTITUTE

BEYOND TRANSPARENCY

PLANNING, ATTAINING, AND USING OPEN ELECTION DATA IN CITIZEN ELECTION MONITORING

BEYOND TRANSPARENCY

PLANNING, ATTAINING, AND USING OPEN ELECTION DATA IN CITIZEN ELECTION MONITORING

NDI
NATIONAL DEMOCRATIC INSTITUTE
FOR INTERNATIONAL AFFAIRS

ISBN 978-1-880134-43-6

TABLE OF CONTENTS

INTRODUCTION TO PLANNING, ATTAINING, AND USING OPEN ELECTION DATA

INTRODUCTION

With advancements in data collection and management technologies and increased efforts by governments across the globe to open data, information about electoral processes has never been more available or accessible. This includes information like voter registration statistics, candidate nomination information, polling station locations, campaign expenditures, and other data typically housed within government institutions. Such **transparency** can help keep citizens and stakeholders informed on election procedures and developments. Open data can also be used to promote **accountability** and improve government performance**.**

The Open Election Data Initiative (OEDI) from NDI offers civil society, election administrators and technologists with the concepts and tools to effectively advocate for, implement, and use election data that is truly open. Over the years NDI has created numerous tools and guidance documents to help stakeholders in this effort. The Unleashing the Potential of Election Data guide lays out the four guiding elements of electoral integrity, which include transparency, accountability, inclusiveness, and competitiveness, and the nine key principles for making election data open. Fully "open" election data is timely, granular, available for free on the internet, complete and in bulk, analyzable, non-proprietary, non-discriminatory, license-free, and permanently available. The Election Data Guide also details the key categories of election data that are being collected and produced during the pre-election, election day, and the post-election period. The Latin America Data Inventory applies open data principles to elections in countries in Latin America and the Caribbean where electoral management bodies (EMBs) and governments took steps to "open up" their data, and illustrates how open data principles translate into practice. The OEDI Election Data Academy provides users with hands-on modules and exercises on how to analyze real election datasets, such as polling station lists, ballot qualification data, and voter registration data, once they have been obtained in an "open" format.

The *Beyond Transparency: Planning, Attaining, and Using Open Election Data in Citizen Election Monitoring* tool seeks to support citizen monitors to integrate an open election data strategy into their broader observation effort. It outlines specific strategies that observer groups can employ for open election data collection, analysis, and advocacy and highlights various challenges that monitors may

face along the way. The tool is made up of six steps that provide strategies for implementing an open election data plan according to the specific country and electoral context and monitors' capacities. The first step defines the overall observation strategy by helping users define their priorities and objectives for election monitoring, which will structure the strategy for open election data plans. Next, users will think through what data is available, where to find it, and how it may look. In the third step, users will develop and finalize a data analysis strategy tailored to their priorities, capacities, and timelines. In the last three steps, users will work through key considerations for obtaining election data, including methods for requesting data from electoral management bodies, and consider the ways users can monitor success of their open data strategy.

WHY SHOULD CITIZEN ELECTION MONITORS USE OPEN DATA?

As trusted and independent actors, nonpartisan citizen election monitoring groups are particularly well-suited to collect and scrutinize open election data, especially when data may otherwise be difficult for the public to digest, aggregate, or interpret. Open election data can support broad citizen monitoring objectives by **supplementing and expanding upon traditional observation findings**, and **ensuring open election data is understood in a credible and impartial way**.

Open election data is comprehensive, external information that can be gathered without physical observation, thereby providing monitors with additional context that can sharpen their analysis and bolster their observation efforts at minimal costs. Open data can benefit multiple facets of citizen election monitoring because:

- **Open data is official data.** By using electoral data that originates from official government sources, civil society can operate with the same assumptions and facts as stakeholders and institutions in the process. In addition, using data provided by government bodies creates an accountability link to those institutions and protects monitoring groups from accusations of anecdotal or inaccurate information.

- **Using open data is generally not resource intensive.** Open election data offers monitoring organizations data sets that they do not have to physically collect through direct observation. While collecting and analyzing open data may take some staff time, and in some cases, data analysis software, it typically does not require the large-scale training and deployment of observers, thus minimizing the amount of resources needed for the effort. Groups with resource or funding limitations can use open election data to participate in aspects of the electoral process they otherwise may not address. Open election data is also a cost-efficient way to bolster more traditional observation efforts by adding additional layers of analysis and sophistication.

- **Open data can promote greater engagement throughout the electoral cycle.** Open election data is generated throughout the electoral cycle which means that groups can be active early in the electoral process. Even if formal observer recruitment, training and deployment for an election day observation effort is unlikely to occur until closer to election day, core staff can be collecting and monitoring open election data months and even years in advance as part of comprehensive election monitoring in the pre-election and inter-election periods. Providing such analysis can help establish or boost an observer group's reputation as a credible voice citizens can turn to for information.

- **Open election data can support and contextualize other election observation exercises.** Election observation activities, particularly those that involve the training and deployment of observers, require information and planning. Some open election data, such as polling station lists or voter registration information, can be critical to these exercises. By obtaining this information—and conducting thorough analysis of any trends or deficiencies within the data—as early as possible, citizen election monitors will be better positioned to develop deployment plans and observation forms in a timely fashion. In addition, information like polling station and precinct data can be built into election day reporting databases early, improving analysis and easing staff workload closer to election day.

By using open data, citizen election monitors can distill large datasets for easier public consumption, identify additional trends in electoral participation and administration, and provide more informed recommendations for election management bodies (EMBs) and other government institutions and stakeholders. However, acquiring and analyzing open election data can be challenging, and it's important that monitoring organizations have a strategy in place to obtain and use open data that best fits their monitoring objectives.

HOW CAN YOU MAKE AN OPEN ELECTION DATA MONITORING STRATEGY?

The amount of open data available for analysis in an election cycle can be overwhelming, and some may not ultimately be relevant to electoral integrity or to a group's specific observation effort. In some contexts, there may be a very limited amount of open data available for use. It's important for monitoring groups to determine how the use of open data can best complement existing observation plans, and develop clear targets for acquiring and using data.

OPEN ELECTION DATA MONITORING AND ADVOCACY TOOL

Keeping in mind the Open Election Data principles, this tool should provide a framework for citizen monitors to integrate an open election data strategy into their broader observation plan and help them confront challenges to open election data collection, analysis, and advocacy.

The tool will highlight the planning stages of developing a logical open election data monitoring strategy that can amplify an observation effort. In particular, it will walk groups through a six step process of developing and implementing an open election data plan that complements their observation objectives, including:

1) **Defining the overall observation strategy**, including goals, priorities, activities and a timeline,
2) **Assessing the availability of open data,** including what data is feasible and viable, and where and when to find it,
3) **Developing an analysis strategy and finalizing an observation plan** with clear and realistic priorities, targets and timelines,
4) **Advocating for open data** from state actors,
5) **Being a good consumer of open data**, and
6) **Measuring success.**

STEP 1

DEFINE THE OVERALL OBSERVATION STRATEGY

At an Open Election Data Academy in 2019, participants discuss how to present data in a way that is informative and best suited to their objectives.

This section should help observers determine their overall election monitoring objectives if they have not done so already. Once observers are clear what parts of the process are most crucial to their monitoring effort, only then can they start to consider how to incorporate open data into such an effort. Observer organizations should have a well-defined overall strategy *first* before pursuing open election data collection and analysis.

WHAT ARE THE OBSERVATION PRIORITIES? WHAT ELECTORAL PROCESSES ARE MOST AT-RISK?

No monitoring group can do everything. Organizations may have to limit their focus based on priority threats to electoral credibility, as well as institutional expertise, timing, and resource constraints. By identifying priorities, monitoring organizations can ensure that the most relevant electoral issues are scrutinized and that resources are allocated in the most effective way.

Groups should determine which parts of the process are likely to have the greatest risk to the overall integrity of the elections. This includes considering risks to competition or a level playing field, risks to participation, risks to inclusion, risks to transparency, and any other challenges that could undermine the credibility of the electoral outcome. This evaluation can include asking:

- What problems have occurred in previous elections? What kind of electoral complaints, if any, were filed or reported?
- What is the political climate for the upcoming elections?
- What, if any, challenges do groups anticipate in the pre-election period, on election day, or in the post-election context?
    - Of these, how likely are these challenges to occur?
    - Of these, are there areas that you are particularly qualified to engage in? That you've observed previously or have specialized knowledge of?
    - Of these, what problems could be most effectively mitigated by election monitoring?
- What is the international, citizen, and political party election monitoring landscape? Are there any other organizations that are already covering some of these topics?
- Are there any rumors or hearsay about the upcoming elections? Are there particular topics the public is concerned about? Issues related to election rigging that are commonly referred to?
- Are there any changes in the legal framework, administrative procedures, power dynamics, and public opinion that could influence how the electoral process occurs?

After asking the above questions, consider how problems in accountability, competition, inclusion, and transparency may actually appear during the electoral process and how those issues could undermine electoral integrity. Complete Exercise A: Risk Diagnostic and Prioritization.

EXERCISE A: RISK DIAGNOSTIC AND PRIORITIZATION

After asking yourself questions about electoral risks, consider how problems in accountability, competition, inclusion, and transparency may actually manifest themselves during the electoral process and how those issues could ultimately undermine electoral integrity. For example, mistakes or disorganization in the ballot qualification process would be problematic because they could create advantages or disadvantages for a certain party or candidate, disrupt the right of citizens to run for office, limit voter choice, and/or harm public confidence in the process.

Brainstorm what problems the group is concerned about, and how those pose a risk to the elections. See where these problems overlap with the open election data process categories. Then determine, through a 1 to 5 scale, what the *likelihood* is of such an issue occurring in the elections and how *severe* the impact may be. Then, by comparing both the likelihood of occurrence and severity of impact, prioritize each issue with its corresponding open data category.

TABLE A: RISK DIAGNOSTIC AND PRIORITIZATION

Open Election Data Process Category	Risk *What is the challenge you've identified?*	Likelihood of Occurring	Severity of Impact	Priority *Where does this rank in your observation priorities?*
Legal Framework				
Electoral Boundaries				
Election Management Body and Administration				
Election Management Body Processes				

Open Election Data Process Category	Risk *What is the challenge you've identified?*	Likelihood of Occurring	Severity of Impact	Priority *Where does this rank in your observation priorities?*
Election Security				
Political Party Registration				
Ballot Qualification				
Election Campaign				
Campaign Finance				
Voter Registration				
Voter Lists				
Voter Education				
Polling Stations				
Election Results				
Electronic Voting and Counting				
Electoral Complaints, Disputes and Resolution				

WHAT IS YOUR OBSERVATION STRATEGY?

Groups should reconcile their overarching election observation goals with the major priorities identified. Some questions to consider when defining your observation strategy include:

What observation methodologies suit your goals?

- What major components and activities are necessary to achieve these goals and monitor these processes?
- When will the organization—at least core staff—be engaged in the electoral process?
- What processes should you observe and to what extent? Which aspects will require the training and deployment of observers?
- What kinds of funding or personnel limitations exist and how will you work within these constraints?

Taking into consideration the group's broad observation goals, the initial risk diagnostic and prioritization, and observation strategy, groups should review the previous chart and reduce it to only the top eight or ten issues. Does anything need to be reorganized when you narrow the priorities?

Keep in mind that some categories are universally key to electoral integrity, regardless of local context. That includes **Election Management Body Processes** (how the elections are administered), **Ballot Qualification** (how political parties and candidates pass eligibility requirements to be on a ballot), **Voter Registration** (determining who in the population can cast a ballot), and the **Results and Complaints** processes (determining whether the will of the voters is reflected in the electoral outcome). Groups should always consider prioritizing these categories highly.

WHAT ARE THE OBSERVATION OBJECTIVES?

Election monitoring can accomplish many goals, such as **increasing public confidence, enhancing citizen engagement, deterring problems and exposing them when they occur**, and **providing important unbiased assessments on the integrity of the elections** to the public. There are a number of questions to consider when clarifying your observation goals, such as:

- Why is election monitoring important to this election? What aspect or aspects of electoral integrity needs this kind of intervention?
- What type or aspects of election monitoring might be the best intervention? The mass mobilization of observers? The collection, analysis, and dissemination of accurate, evidence-based data?
- How do you want to build on your previous monitoring efforts?

After this brainstorming, monitoring groups should think **pragmatically**. What can you realistically achieve in the time frame and resources that you have? Try to narrow your objectives to one up to five. Monitoring groups should be able to succinctly describe their observation objectives in one or two sentences or bullet points.

WHAT IS YOUR OBSERVATION STRATEGY?

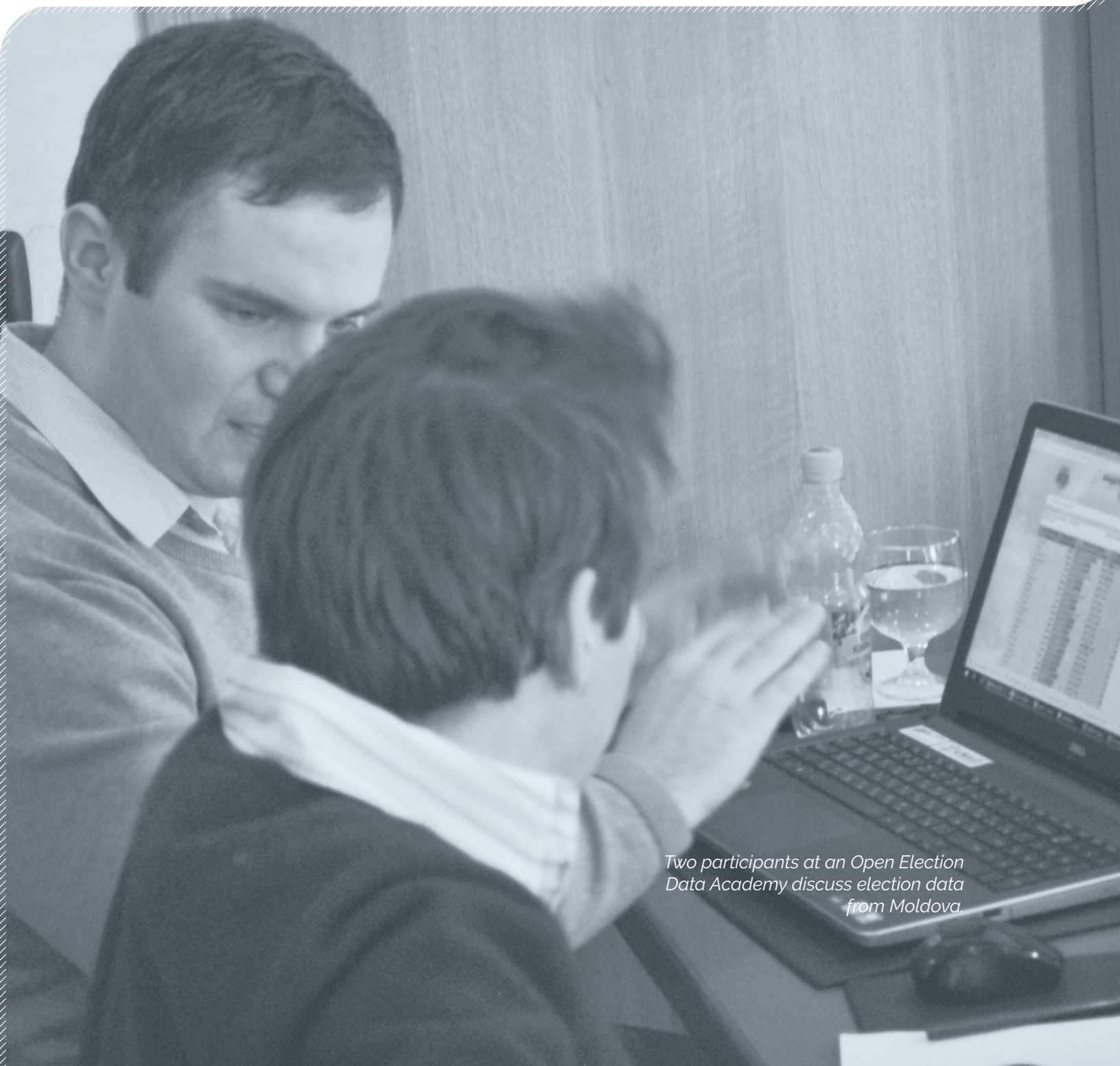
Groups should reconcile their overarching election observation goals with the major priorities identified. Some questions to consider when defining your observation strategy include:

- What observation methodologies suit your goals?
- What major components and activities are necessary to achieve these goals and monitor these processes?
- When will the organization—at least core staff—be engaged in the electoral process?
- What processes should you observe and to what extent? Which aspects will require the training and deployment of observers?
- What kinds of funding or personnel limitations exist and how will you work within these constraints?

Keeping in mind the top priorities defined in Exercise A, groups should consider how open election data can help to achieve their observations goals. Specifically, how would open election data monitoring fit with the objectives? The next step of the *Beyond Transparency: Planning, Attaining, and Using Open Election Data in Citizen Election Monitoring* tool will walk groups through how to assess the open election data environment.

STEP 2

ASSESSING THE AVAILABILITY OF OPEN ELECTION DATA



Two participants at an Open Election Data Academy discuss election data from Moldova.

Understanding when and where to collect open election data will help inform observation plans. Groups should conduct a preliminary assessment of the open election data environment, including what data is likely to be feasible or viable, where it resides, and when it may be available. This assessment should also help organizations determine what datasets may be associated with the election procedures and processes they've identified as high priority.

ELECTORAL PROCESSES ASSOCIATED WITH OPEN ELECTION DATA

The most valuable information regarding election day processes and environment is still the information that well-trained and deployed election monitors can collect about the integrity of the voting, counting, and tabulation processes. However, there are a number of aspects of the electoral process throughout the pre- and post-election periods that may have accompanying open data that can enhance observation findings. Groups should familiarize themselves with these key categories at openelectiondata.net if they have any questions about what they entail.

OPEN ELECTION DATA: FEASIBILITY AND VIABILITY

There will be limits to what election data actually exists, what data is available to the public, and whether that data is useful or usable. In order to maximize time and resources, monitoring groups should only attempt to pursue election data that is both **feasible** and **viable.** Feasibility means that monitoring groups can *reasonably* access the data, or reasonably expect the government to have such data. Viability means that the data is both relevant to your electoral analysis and can accurately be used for analysis.

Open election data is *not* feasible if:

- **The data isn't collected in the first place**. For instance, monitoring organizations may want to use ballot qualification data to determine how many candidates from a certain ethnic or religious background were accepted, however if that information is not captured in candidate nomination forms, then it likely doesn't exist. If data is not being collected and should, then such a dataset can feed into your advocacy plan (covered in Step 4).

- **The data exists but is not available in the timeframe needed.** For instance, if a group wants comprehensive information on electoral disputes a week after the elections but the dispute resolution process is ongoing, then that data is not feasible at that time.

- **The data is legally protected from the public.** Some data, or some fields in a dataset, may be protected by law. This is often the case for personally identifiable data—such as national

ID numbers—or for protected individuals such as witnesses, victims of abuse, some security personnel, etc. However, data produced and collected by EMBs is collected and paid for by citizens and should inherently be public data. Additionally, this data is often available publicly during voter registration and election day even if in limited scope. Just as elections and government belong to the people, public data belongs to the people.

- **Other barriers exist that make the data impossible to reasonably acquire.** For instance, if fees associated with accessing the information are prohibitively expensive (which also violates basic principles for open election data).

Open election data is *not* viable if:

- **The data is incomplete.** While missing data may prevent groups from conducting comprehensive analysis, groups will have to determine what, if any, analysis can occur with the limited data. For example, completeness of data is especially important when there's a geographic component to the data because missing data for a particular region, state, or district can give the impression that an EMB is biased against a particular party or candidate.

- **The data is inaccurate or of such poor quality or provenance that it cannot be trusted.** Groups will likely struggle in conducting analysis with data that is inaccurate. There are situations where names and other information is no longer accurate due to digitizing and transcribing processes such as those from non-Latin characters. There are some instances where how the data was collected may be unclear or not well documented. In addition, if the dataset is perceived to be particularly sensitive and the exact process of how it was collected is unclear or inconsistent, then it may not be viable.

- **There is no government consensus regarding what data is "official."** In some cases, data from different government sources may contradict one another or otherwise not be considered formal or official. If there is no clarity regarding which data holds precedent or is officially used by the government or EMB, then it may not be viable. However, groups should still try to attain as much of this information as possible because comparing multiple datasets can provide insights into the data quality.

- **It is impossible to receive the data in a format that can be analyzed within the necessary timeframe.** This is particularly important when working with files that contain large datasets that must be converted to a machine readable format. There are some formats that may take time or special expertise to convert (such as PDFs) but can be managed with the right resources. However, the process for converting other file formats, such as image files, can be very long and laborious and put groups at risk of not being able to conduct timely analysis.

- **The data does not enhance your analysis.** Just because some open data exists doesn't mean that it will lead to fruitful analysis regarding electoral integrity. Always consider the goals of analysis before wasting time acquiring data that will not ultimately be helpful.

Groups should take the time to assess what desired data is actually *feasible* and *viable* as they develop their open election data strategy.

SOURCES

Generally speaking, most open election data will reside within the **election management body (EMB)**. This includes items such as voter registration lists, candidate or party registration information, polling station lists, campaign finance reports, and most other administrative election data. However, there is a chance some data of interest may exist with other government or semi-government institutions. For instance:

- **Legislative bodies/executive offices –** Lawmaking and budgetary bodies may contain valuable data related to the most consolidated and up-to-date legal framework for elections, as well as some budget or procurement data relevant to the electoral process.

- **Judicial bodies –** Election tribunals, appeals courts, and regular judiciary bodies will likely be important for election complaint and dispute resolution data, and may also be relevant to other election charges (for instance, in ballot qualification appeals or challenges to aspects of the legal framework or election boundaries).

Open-Source Spatial Tools

Geopolitical information, such as information on administrative boundaries (like prefectures, counties, and states) and electoral boundaries (electoral districts with associated elected representatives) should be retrieved from official government sources. However, other helpful analytical tools, such as GPS coordinates, shape files, and mapping software, can often be obtained for free from other sources that are open-source, proprietary, and cloud-based. For example, adding GPS coordinates to voter registration centers may help identify geographical gaps and be correlated with lower rates of registration in some districts.

- **Census/statistics bureaus –** Departments that work with population data, such as census or statistics bureaus, can provide information on vital statistics, population, and demographic information.

- **Interior or land ministries/geological surveys –** Departments that deal with geopolitical and spatial information, such as an interior or land ministry, may have information, like official maps, shape files, or GPS data, that groups may consider using in their analysis.

- **Peace and security actors –** Security forces may have information regarding electoral violence incidents, campaign or election violations, or other information related to electoral security and criminal election offenses.

- **Third party vendors –** Occasionally governments and EMBs may subcontract out management of some election data to third party vendors. While governments should, in theory, be able to access and disseminate this data from their contractors at any time, there may be situations where it is easier to work directly with vendors to acquire data.

Every government is different so keep in mind any other local or national institutions that could have relevant information. For instance, in some places voter registries are linked to civil registries – like national ID services – and those bodies may contain significant data relevant to a group's needs.

TIMING

Not all open data will be available at the same time, and the electoral calendar established by the EMB will have a substantial role to play in when organizations can reasonably expect data to be released. For instance, it is unreasonable to expect the EMB to have a polling station list available a year before election day. Polling station lists are often drawn up or modified following voter registration exercises to ensure polling stations are well distributed among the voting population.

Some data may be relatively **static** (remains unchanged after it is generated) throughout the electoral process, such as census data or spatial data. Therefore some static data should be able to be acquired at any point. In some cases, groups may even be able to use open election data to provide some analysis on parts of the electoral process that they may have otherwise missed. For instance, electoral boundaries may be drawn several years in advance of an election when a group may not yet be ready to conduct analysis. However, because that information is likely to remain static after it is generated, groups can use that information and comment on the process whenever they are ready to conduct pre-election analysis.

Groups should consider and plan on when other electoral data is likely to be generated. In addition, groups should consider whether the data is likely to be complete and static after generation (such as a final candidate list following ballot qualification) or **variable** and requiring regular collection (such as reports of electoral violence as they occur).

ILLUSTRATIVE DATASETS, SOURCES, AND TIMING CONSIDERATIONS

Groups should always consider the broad range of data available to them that can support their observation priorities. ***See Illustrative Datasets, Sources and Timing Considerations in the Supplementary Booklet*** for the key electoral categories associated with open election data, with datasets that *may* be relevant to them. This list is illustrative, as many datasets and sources will be unique to the country or political context.

VERSATILE DATASETS

As you can see from the table, some data sources can be used for analysis of multiple electoral processes. Such datasets may be considered a high priority to capture because they serve multiple purposes. For instance, a set of consolidated election laws and amendments and EMB regulations can help the analysis of all major parts of the electoral process. Population and demographic data—such as a census, demographic or social surveys, development reports, etc.—can provide important context and baselines for assessing access and participation issues. Voter registration lists and statistics can be used to analyze several electoral processes. In addition, most of these datasets are typically publicly available and don't tend to change after they're produced. The chart below highlights some of the more common sets of data that can support analysis in multiple electoral categories.

TABLE 1: VERSATILE DATASETS

Dataset	Relevant Election Processes for Analysis	
Consolidated electoral laws and amendments, and EMB regulations	Legal framework EMB administration Election security Ballot qualification Campaign finance Voter list Polling stations Electronic voting/counting	Electoral boundaries EMB processes Political party registration Election campaign Voter registration Voter education Election results
Population and demographic data	Electoral boundaries Voter registration Voter list Voter education Election results	
Voter registration list and statistics	Electoral boundaries Voter registration Voter list Voter education Election results	
Maps and spatial data	Electoral boundaries Voter registration Polling stations Election results	
EMB procurement data and budget	EMB processes EMB administration Electronic voting/counting	

MATCHING OBSERVATION PRIORITIES WITH OPEN DATA AVAILABILITY

Open election data should ideally meet the nine principles for open data, including being available for free on the internet. In these cases, groups should be able to access data easily and relatively effortlessly. However, sometimes data does not meet all of these criteria. For instance, data may not be available in as timely a manner as it should, may not include the level of granular detail necessary,

or may be in a format that is not computer analyzable such as a PDF. While reviewing each election data priority, groups should consider:

- Is this data both feasible and viable?
- What data is immediately available and accessible, such as available for free on the internet? Or, if it's too early in the process, what data do you believe will be easily available and accessible?
- Can you use the current format available to conduct your analysis, or will the format need to be changed or converted?
- Is it available at the smallest (i.e., granular) level needed for substantive analysis?
- What, if any, datasets will you need to request from government institutions? What, if any, alterations to pre-existing datasets – such as formatting changes – will you need to request from government institutions?

Take only the top eight election categories as identified by the risk assessment and review the **feasibility, viability and compliance with open data principles** for each category. Answer the questions included in the assessment forms (included in the Supplementary Booklet) for each category.

Open Data and Cross-Cutting Issues

While groups should try to prioritize their observation and data collection plan based on major election functions, it's important to keep in mind cross-cutting issues that are relevant throughout the electoral process, especially the **participation of women, youth, minorities, and vulnerable groups.**

The use of open data can be particularly helpful in collecting information regarding women and other marginalized populations in many key election procedures. For instance, voter registration and turnout data, candidate information, and EMB staff profiles can and should be disaggregated by gender and other demographic information when possible. In addition, the distribution of polling stations, electoral boundaries, or security forces can be overlaid with demographic information to identify any trends that could impact certain populations.

STEP 3

DEVELOP AN ANALYSIS STRATEGY AND FINALIZE YOUR OBSERVATION PLAN

At an Open Election Data Academy in 2019, participants think of ways to visualize data in a graph.

The analysis of open data should not supplant or replace other observation activities. Rather, open data analysis can help enhance monitoring strategies and provide greater context and support to monitoring conclusions. Just because an organization has decided to use open data to analyze a part of the electoral process does not mean that it should not also want to deploy observers to collect real time information on the process.

Practicing Data Analysis

The Open Election Data Academy includes practical modules for readers to learn how to analyze common datasets released during elections. The modules guide readers through how to make the data work for you.

OPEN ELECTION DATA MONITORING AND ANALYSIS PLAN

A data analysis plan is a detailed document outlining procedures for conducting an analysis on open election data and what various steps in the analysis are going to be. An analysis plan should be created prior to beginning data analysis in order to guide how a group is going to collect and analyze the data. This document is a living document which may be revised during the data analysis. The following table outlines key components and example inputs of a data analysis plan. Groups will need to consider availability of data, capacities of staff, resources needed, etc. The remainder of Step 3 provides more details and exercises to help groups think about and complete a data analysis plan.

TABLE 2: SAMPLE DATA MONITORING AND ANALYSIS PLAN

Problems					
Objectives					
Research Questions	**Data sources**	**Type of analysis, variables and subgroups**	**Staffing and resources**	**Notes**	**Timeline**
What are the rates of voter registration across geographic divisions and genders?	Voter registration dataset from election commission	Quantitative analysis of changes in voter registration rates across geographic regions through by gender	Elections Program Manager; Data Analyst; Data Processing Programs (e.g., Excel or google sheets if small dataset)	Voter data from EMB sent as a .csv file. Population data from Census Bureau. Missing data points recorded as N/A.	Oct 7-14: gather data, create guidelines for storing files, begin documentation. Oct 15-25: clean data, update documentation. Oct 26-Nov 2: conduct quantitative analysis, update documentation.

MATCHING OPEN DATA WITH ANALYSIS CAPACITY

A key consideration of a group's data analysis plan should be internal capacity to conduct data analysis. Depending on what groups plan on analyzing and the size of the dataset, specialized knowledge may be required. A Data Science consultant may need to be hired to help with either merging multiple datasets or summarizing the data at a higher level (making the dataset smaller) so more can be done by in-house staff. Or the consultant may need to conduct all of the analysis for large datasets.

Q: What is a *large* dataset?
A: More than 1 million records

Many organizations use Microsoft Excel or Google Sheets for data analysis. However, these programs have some limitations. At the time of publication, an Excel file cannot have more than 1,048,576 rows/records. And the total number of characters in a cell is capped at 32,767.
Microsoft documentation of limits: https://bit.ly/excel-limits
Google documentation: https://bit.ly/sheets-limits

For large datasets like voter lists, free and open-source programming languages such as R or Python can be used. However, they do require more specialized knowledge of programming and the relevant data analysis modules/packages.

Analysis of open data will often require basic arithmetic to identify trends and draw conclusions, including identifying percentages of increase and decrease, averages, standard deviations, changes over time, and proportional comparisons. In some cases more advanced analysis techniques may be required for more sophisticated analysis (such as calculating correlations between variables or running machine learning models). A group's access to data management and analysis resources and comfort with statistics and programming could impact how deeply they can analyze certain datasets on their own.

Groups should reflect what capacity they have to:

- Retrieve datasets
- Collect, 'scrape', or convert datasets if necessary
- Clean and parse data
- Summarize datasets, especially if they are large (i.e., more than 1 million records)
- Repeat steps above if necessary
- Analyze complex datasets or merge simple datasets into a complex dataset
- Work with spatial data and/or mapping software or packages (e.g., Q-GIS or relevant programming libraries like Folium, GeoPy, and/or Leaflet)
- Conduct advanced analysis (e.g., correlations or machine-learning models)
- Visualize data
- Storytelling with data

Groups limited in some capacity can consider if they have the budget to work with a Data Science consultant to provide assistance. Organizations can test out their analysis skills by participating in modules at the Election Data Academy online and should answer the questions in Exercise B: Organizational Capacity to Use Open Election Data.

EXERCISE B:
CONSIDER ORGANIZATIONAL CAPACITY TO USE OPEN DATA

Groups interested in collecting and analyzing open election data should consider the following questions to determine how in-depth their effort should be.

1) Has your organization previously analyzed open election data, such as voter lists or polling station distribution? If so, how?

2) Does your organization understand the difference between good, reliable data and bad, unreliable, or incomplete data?

3) Has your organization used data quality checks in the past before conducting analysis of observation findings?

4) What, if any, database software is your organization accustomed to using?

5) Does your organization have the ability to convert open data into a usable format?

6) Does your organization have any experience with mapping software? If so, what?

7) Does your organization know how to import, clean, and manipulate data in spreadsheet software?

8) Is someone within the organization available to use and maintain a database throughout the electoral cycle?

9) Does your organization use basic principles of statistical sampling and inferential statistics?

WHEN DATA IS EASILY RETRIEVED

As noted, the **retrievability** of data will impact the type and level of data expertise groups may need. The ease or difficulty in retrieving data needs **to be examined** as a key part of the data analysis plan. Ideally, most, if not all, the election data will be available, for free, on the internet. In this case, groups should be expected to download and import data into a CSV, XML, or JSON file (this is usually the case for large datasets and quantitative information), or download and conduct some manually data input or qualitative analysis (this could be the case for forms, rules and regulations, security reports, notifications, and data that does not lend itself to consolidation). However, in some instances data may not be available online or according to key open election data principles. In situations where data is not easily retrieved, groups should consider other strategies, such as "scraping" data to collect and convert the needed information as well as conducting advocacy, the latter of which is detailed in Step 4.

Technical issues to keep in mind when requesting data

When requesting a dataset, it will be important to understand what parameters were used for each of the following so that the data can be properly read into your analysis software:

1) **Field Separators (delimiters):** A tab-delimited file or comma-separated value (CSV) file are text format files. In order to tell the program where each field begins and ends in an imported or exported file, the fields must be separated (i.e., delimited) with a character such as a comma (,). This character is called the field separator or delimiter. Tab is another common delimiter.

2) **Text-Qualifier:** If a field contains the delimiter character within its text (e.g, an address or name in a voter list dataset), the program will interpret the comma as the end of a field rather than just part of the text. In order to prevent this mistake, fields must have a qualifier. The qualifier is put around each field to note that all the text inside is the same field even if it contains a comma. The most common qualifier is double quotes ("). Most programs will include a qualifier when saving in CSV format.

3) **Character Encoding:** When files are saved, a specific character encoding is used. UTF-8 (unicode) is the most common. However, due to different lan guage and scripts, a different kind of encoding might be used.

DATA PREPARATION

Another consideration when completing a data analysis plan is to think through how much data preparation will be needed ahead of beginning data analysis. **Data preparation** is the process of cleaning, structuring, and enriching raw data into a desired output for analysis. Data preparation takes the highest amount of time and resources. However, your analysis will only yield meaningful results if your data has high quality. It's important to ensure that your datasets achieve data integrity, in terms of their completeness, accuracy, and consistency. See Step 5 for more details on best practices in data preparation.

CHOOSING TYPE OF DATA ANALYSIS

Groups should be creative regarding what open election data can tell them about the integrity of the electoral process. Depending on what data sources and processes organizations are examining, open data analysis can provide insights regarding the **competitiveness**, **inclusiveness**, **transparency**, **efficiency**, and **credibility** of the electoral process. Monitoring organizations can draw meaningful conclusions from open election data through the following strategies:

- **Summarizing Data –** Consolidating and summarizing complex datasets or otherwise disaggregated information can help clarify the outcome of certain election procedures, and can demonstrate where deficiencies in the process – or in the data – exist. Module 3 of the Election Data Academy demonstrates how to find summary statistics for a particular variable in a dataset to gain a clear understanding of data at a glance.

- **Qualitative Assessment –** There are a number of qualitative indicators that may be present in open election data. Monitoring groups can use comparative information and democratic standards to help distill and present qualitative conclusions, particularly for data that cannot be quantified (such as legal document analysis). Where possible, however, groups should try to quantify qualitative information that can be gathered from open data – for instance, through percentages, averages, changes-over-time, and numeric comparisons.

- **Quantitative Assessments –** Applying quantitative analysis to open election data can provide the most robust and detailed findings. In some cases, monitoring organizations can combine several datasets to develop a multi-faceted quantitative analysis. For example, combining voter registration data with census data provides insights on the eligible voting population as demonstrated in Module 7 of the Election Data Academy.

"Scraping" Data

Data scraping is a method of collecting and converting data from sources that are not machine analyzable format but otherwise exists, such as in a PDF or in a table displayed on a webpage. The process of data scraping and validation can be time-consuming and usually involves some form of software. Some software does not require any knowledge of coding (e.g, Dexi.io). However, if the data is considered valuable to the observation analysis, such as results by polling station level or voter list information, groups should consider investing in data scraping tools and time allocations in order to collect and convert the data.

Some data will lend itself more towards qualitative or quantitative analysis, but groups should strive to conduct the most in-depth analysis they can with the data available, applying as many of these strategies as possible for each dataset. Citizen monitors should review this section of the Election Data Guide to get more information about what kinds of analysis can be conducted for each election category. Groups should also take advantage of the Data Academy modules available at openelectiondata.net. For additional details on how to analyze particular types of election data, partners can review NDI manuals such as Building Confidence in the Voter Registration Process and The Quick Count and Election Observation.

FINALIZE YOUR PLAN

At this point, monitoring organizations should have identified what election processes are a priority to gather open election data for, and whether this data is feasible and viable. Groups should also have conducted a preliminary assessment regarding:

- What desired data can be accessed immediately in a sufficient format and level of granularity;
- What data will need to be requested and in what format and level of granularity;
- What data may need to be 'scraped' or converted into a different format;
- Where the data resides;
- When the data will be available; and
- The kind of technological and analytical needs associated with this data.

Based on this information, along with an organization's monitoring strategy, groups should parse out what data can (and should) be gathered by observers, and what data can be gathered via open official sources. You should be ready to fill out Exercise C – Datasets, Sources, Timing, and Requests.

Different Types of Data

The term "data" can be used to describe any facts related to the electoral process that are aggregated for the purposes of reference and analysis. Some of this will be statistics, numbers, or rates, while others will be words – such as candidate nomination forms, electoral violence reports, and regulatory frameworks. Groups should be equipped to conduct analysis of many different types of data.

Then, based on your monitoring priorities and objectives, groups should specify a list of observation activities with a corresponding realistic, if not broad, timeline. Plot this out in Exercise D – Open Data and Election Observation Activity Timeline. You can use the chart in Exercise D as a guide, which includes some placeholder information as an example.

EXERCISE C: DATASETS, SOURCES, TIMING, AND REQUESTS

Based on the prioritized election categories and organizational capacity, identify the datasets groups want to pursue, where the data are likely to reside (data source), when data are likely to be available, what format the data is likely to be, and whether groups will need to make any special requests to access the dataset. Review Illustrative Datasets, Sources and Timing in the Supplementary Booklet as needed to ensure you consider your options thoroughly.

TABLE C: DATASETS, SOURCES, TIMING, AND REQUESTS

Election Process	Dataset(s) Needed	Source(s)	What format will it be in?	Will a reformatted request be needed? Y/N	Will a request to access the data at all be needed? Y/N	When is the data available?

EXERCISE D:
OPEN DATA AND ELECTION OBSERVATION TIMELINE

Based on your monitoring priorities and objectives, groups should specify a list of observation activities with a corresponding realistic, if not broad, timeline. You can use the chart below as a guide, which includes some dummy information in it as an example.

TABLE D: OPEN DATA AND ELECTION OBSERVATION TIMELINE

Timing	Process	Observation and/or Open Data Activity
March	Legal Framework	Core team analyzes legal framework for upcoming elections
April	Voter Registration	Core team requests preliminary voter list
June–July	Voter Registration	Training and deployment of voter registration observers
August	Voter Registration	Core team analyzes preliminary voter list
September	Ballot Qualification	Training and deployment of candidate nomination observers
September	Ballot Qualification	Core team requests candidate nomination applications and analyzes them

STEP 4

ADVOCATING FOR OPEN ELECTION DATA

Participants at an Open Election Data Academy listen to a presentation on how to use election data as part of their domestic election monitoring work.

After developing an open election data plan, groups should have a strong idea of what data they can retrieve easily without any specific changes and what data they may need to request or seek modifications for. Both will involve some work from civic organizations, although the latter more so.

WHEN SPECIFIC DATA OR DATA FORMATS MUST BE REQUESTED

In cases where complete and analyzable datasets are not available online, groups will need to take additional steps to acquire the information needed. Monitoring organizations should be deliberate in how and when they make open data requests to ensure they get the data they need. In this effort, groups will need to advocate directly with stakeholders in order to gain access to this data or specific formats. Advocacy is also not a one-size-fits-all strategy – in the same way that stakeholders will have different responsibilities and mandates over election data, advocacy efforts need to be tailored and responsive to the political context and interests of different stakeholders.

Be strategic

Monitoring organizations should understand which data is the most important (the highest priority) and which data they should ask for first. At least then groups can have the best chance of getting the best data, and then can push for lower priority data needs if the opportunity arises. Prioritizing data requests should consider:

- What open data a group absolutely needs within their observation strategy;
- What open data a group needs first to begin observation activities;
- What data is likely to take the most time to acquire given local processes; and
- What datasets, such as some of the versatile sets, will be useful throughout an observation.

Be diplomatic and inclusive

Consider working alongside the EMB and other civic actors to achieve open data principles. Monitoring groups should seek to meet with EMBs early on with their election data needs so that EMBs understand what expectations exist. EMBs may not realize why a certain data format may be problematic or what kind of timeline civic groups operate under. With enough notice, EMBs may even be able to alter data collection and analysis procedures to ensure the appropriate data is available. For instance, EMBs could update an administrative form to incorporate a data field organizations think would be helpful to evaluate.

In addition, other civil society actors and political parties may be interested in the same or similar data. Monitoring organizations should encourage the development of a broad open data community. Civic groups can learn from one another regarding obtaining and using open data, even in a non-electoral

context. Bringing a broad coalition of interested actors to the table may not only help apply pressure on institutions to open election data, but can also maximize transparency and lay a foundation for other open governance data initiatives.

Address the right institutions and people

As previously mentioned, open data may reside in any number of government and semi-government institutions. In order to make sure your request is processed quickly, groups should make targeted data requests to the appropriate office or department and understand if there are specific or official rules or processes for making such requests. Even within an EMB, there may be data that groups want at the national level, as well as data that may only exist at regional or local levels. If your group is not clear what information resides where, that should be a topic for initial discussions with the EMB.

Assess commitments to transparency, accountability, open data

Opening data is one way public institutions can demonstrate transparency and increase public confidence. Data collected by public institutions is a foundation for decision-making with factual evidence, and opening this data allows stakeholders to assess decisions by institutions. Groups should take time to assess what commitments the government and public institutions have made to increase transparency, such as through law reforms on the accessibility of government data or international initiatives such as the Open Government Partnership.

Give sufficient and appropriate timing for requests

Retrieving open election data may take time, especially if the government has to alter formatting or disaggregate data. Monitoring groups should provide institutions with enough time to reasonably deliver data. In addition, groups should keep in mind that not all data is likely to be available at once. Before making a request, groups should review their observation plan and the electoral calendar to make sure they're asking for the right data at the right time.

Be specific

When requesting data, organizations should be clear about exactly what data they want and when they want it. Groups should provide as much guidance as possible to not only ensure data providers understand what they want, but so they won't have to waste time going back for more or different information. In particular, groups should specify in open data requests:

- Data format (i.e., CSV, XML) and their desired delimiters, separators, and encoding (see the Technical Issues text box in Step 3);
- Data parameters (i.e., dates or timing) – For instance, some datasets may change over time. In those cases, groups should be specific about the time range they are interested in.

- Data details (including desired data fields and level of granularity);
- Data summary in addition to detailed dataset to allow data checking; and
- When they need the data.

Be persuasive when necessary

In some cases, monitoring organizations may have to convince government institutions of the importance of releasing certain data. It's important for organizations to familiarize themselves with any previous commitments their government has made to data transparency. This can include membership in the Open Government Partnership, or participation in certain international and regional human rights mechanisms that compel electoral transparency, such as the International Covenant on Civil and Political Rights (ICCPR) which guarantees the right of citizens to participate in public affairs[1] and to seek, receive and impart information[2]. Groups should also reinforce that opening election data can be beneficial to government institutions by:

- building public confidence in the process;
- enhancing voter education and dissuading misinformation;
- improving the link between citizens and government; and
- demonstrating EMB and government institutions' commitment to transparency and accountability.

In some extreme cases, monitoring organizations may have to apply greater pressure to obtain election data. While placing increased pressure on institutions and stakeholders may feel daunting, groups should consider open election data as a critical aspect to strengthen electoral integrity and government accountability through more transparent elections. Some countries offer legal avenues for groups to gain government information that is otherwise not provided openly. This may require filing a Freedom of Information (FoI) request, or in more severe circumstances, a legal claim. While these approaches can be effective remedies to uncooperative governments withholding data, they can create hostility with government institutions and can take a very long time to process. Groups should only pursue such action if they determine the data to be absolutely crucial to their electoral analysis, and have enough time to undergo a lengthy process without the data becoming irrelevant.

1 [1] ICCPR Article 25: "Every citizen shall have the right and the opportunity, without any of the distinctions mentioned in article 2 and without unreasonable restrictions: (a) To take part in the conduct of public affairs, directly or through freely chosen representatives," http://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx

2 [2] ICCPR, Article 19, section 2: "Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice." http://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx

Consider level of data literacy within ecosystem and among stakeholders

A basic level of data literacy among officials or commissioners will be needed to support innovation and commitments to data transparency. Groups should remember that the officials they will contact may have varying levels of data literacy and thus may not understand the relevance or technical specifications of the data requested. Addressing the right people or institutions and being specific becomes particularly important in these contexts. In some cases, groups may need to provide education to key actors - such as election officials, other CSOs, the media or the public - around data concepts in order to promote an ecosystem of data literacy. This will allow stakeholders to better understand the importance of high-quality data, what useful data should look like and the uses of data.

Balancing data privacy and transparency

Transparency is one of the hallmarks of a credible election and citizens must have access to election-related data. At the same time, privacy is considered a human right in many countries, as well as under several regional conventions. Notably, the European Union's (EU) General Data Protection Regulation (GDPR)3 provides a comprehensive framework for privacy across public and private sectors that may have an impact on privacy protection beyond EU borders. However, privacy rights, like all rights, are not absolute and weigh against other competing public interests and international standards, including the need for transparency.

Most electoral data highlighted in the Open Election Data Initiative does not include individual-level data, and those that do – such as campaign donor information or the voter list – are considered public data and subject to disclosure obligations. However, governments may have legal protections around certain types of data, which may be in place to protect vulnerable populations. Groups will need to be sensitive to the risks that open individual-level data may present to these populations. For instance, in some countries, publicly available voter rolls exclude the information for exceptionally vulnerable citizens, such as victims of domestic violence or people in the witness protection program. In other cases, personally identifiable information in voter data like national ID numbers may be scrubbed from public voter lists. However, these are often made available to observation groups conducting an analysis of the rolls with the agreement to not to publicly disclose the restricted data and to responsibly store the data.

Groups should understand the varying privacy legal frameworks, cultural practices, and risk profiles in their social and political context and be prepared to navigate those constraints. Broadly, however, it is important to emphasize that data collected through public funds and through public institutions is inherently public data that belongs to a country's citizens.

3 The GDPR took effect in the European Union in May 2018. You can find more information on Regulation (EU) 2016/679 (GDPR) at https://gdpr.eu/.

STEP 5

BEING A GOOD CONSUMER OF OPEN DATA

A citizen election observer presents data from Tunisia during an Open Election Data Academy.

Institutions may be more responsive to open election data requests if they trust the information will be used responsibly and analyzed credibly. To be a good open data consumer, observation groups should take the utmost care to ensure the data is collected, analyzed and presented credibly. Monitoring groups will likely need to take preliminary steps at cleaning and organizing datasets and verifying their credibility and usability, as data provided to groups often comes in an unprocessed form. In addition, organizations should consider discarding any data that is irrelevant to their electoral observation. Source data is often not cleaned, organized or structured in the way groups would need it to be to begin data analysis, meaning groups will need to prepare their data before analyzing it.

Data preparation is the process of cleaning, structuring and enriching data into a desired output for analysis. It will often be a time-consuming task, but it is an essential step to ensuring data is ready before analysis and avoid common pitfalls during the data analysis process.

Generally data preparation includes:

- Downloading and/or converting data into machine readable format, if necessary;
- Importing data into data analysis program (e.g., Excel, Google Sheets, or open-source programming languages like R or Python);
- Cleaning data to check for duplicates, missing data, and bad character encodings;
- Structuring data into the right format (numerical, text, etc) and wrangling data to ensure consistent structure across dataset; and
- Checking data for accuracy such as outliers, wrong categorical data, and other inconsistencies.

In addition to the topics covered in this step, groups should consider reviewing the five major steps in the data analysis process in Module 3 of the Election Data Academy.

DOUBLE CHECK SOURCE

Prior to downloading or restructuring datasets, groups should make absolutely sure they are 1) Using the most up-to-date data, 2) Using the official dataset, and 3) Understand how the data was collected. Simple browser searches may pull up unofficial data on certain topics – for instance, surveys conducted by third party sources – or may link users to an old or outdated government source. While these sources may be used to complement and/or confirm the quality of the data groups are provided, unofficial datasets cannot be the principal source of analysis.

RETRIEVING DATA

In an ideal world, all election data would be available, for free, on the internet in a machine readable, non-proprietary format according to Open Election Data principles that would make it easy to

download, import, and analyze data. However, many governments are not at this level of data transparency yet.

Hopefully groups have advocated for data with public institutions that understand and have committed to open data principles, but if not, groups may still be able to gain access to the data that they want. When an EMB or other institution provides data in a non-machine readable format such as a PDF, groups can use software or techniques which convert the PDF file into a CSV or other machine readable format. Sometimes a single dataset may be divided into separate files by geographic or administrative divisions, so a group might need to combine into a single dataset to conduct analysis. Other common problems can be overcomed with a little flexibility and creativity.

PARSING, RESTRUCTURING, AND CLEANING DATA

Even when data is available in the format and detail observer groups would like, staff may still have to sort, organize, and "clean" data in order to optimize it for analysis. Cleaning data (sometimes called "scrubbing") means making any small formatting changes needed to run analysis. For example, before merging two datasets together by the name of geography, you may need to reconcile slight variations in the spelling and/or need to eliminate trailing spaces. Cleaning data may mean isolating or removing data, records, or fields that are corrupt, inaccurate, or superfluous. This is usually only referred to in the context of large datasets with multiple records. Depending on the state of the data and the goal of a group's analysis, data cleaning and organization could take substantial time.

QUALITY CHECK DATA BEFORE ANALYZING

Just because open data comes from government sources does not necessarily mean it will be perfect. As good consumers of open data, citizen monitoring groups should check to see if there are any errors or discrepancies in the data that could indicate a larger flaw in the dataset. By doing these checks prior to analyzing the data, groups can avoid wasting time working from bad data, and won't be at risk of presenting analysis based on flawed information. In particular, groups should conduct preliminary:

- **Accuracy Checks –** Look for any obvious signs of data inaccuracy. Does disaggregated data add up to the aggregated total? If not, why not? Does information at the local, state, and national reconcile? Consider outliers or data that might not make logical sense. For example, if groups are looking at election results data, does the number of invalid ballots exceed the number of votes cast?

- **Consistency Checks –** Look at the consistency of data across sources. For instance, if the EMB has certain data on their website, does that reflect the data they've provided you in a different

format? If there are multiple versions of a dataset, do they all say the same thing? For example, does the polling station list listed on an EMB website have the same number of polling stations as the list available for download?

- **Completeness Checks –** Look for any gaps in the data. Are all localities represented? Political parties? Dates or time ranges? Is there any duplicate data? For example, if groups are looking at a voter list, they would want to ensure that the data from all constituencies are included.

If monitoring organizations discover credibility, consistency, or completeness issues with open data, they should go back to the source provider to clarify where the issues are. It's important for groups to verify that they are working with the most accurate and complete data set. If data problems cannot be clarified or remedied, groups will have to be honest about the limitations in their analysis and presentation of findings. In cases where data is clearly low quality or inaccurate throughout, then that data should be considered unviable and not be used at all.

DRAW RESPONSIBLE CONCLUSIONS AND PRESENT FINDINGS CLEARLY

Groups should understand the limitations of the open data they are working with and treat their analysis and findings appropriately. In particular, groups should be mindful to not misrepresent the data in any way, and make sure that findings are presented clearly so that they cannot be misinterpreted. When drawing conclusions from open data, always keep in mind the electoral integrity context. How severely do your findings impact the integrity of the process? What do they indicate is at risk? For instance, discovering that a list of polling stations is riddled with incorrect addresses is likely more severe than discovering that a list of polling stations has many polling station names misspelled because the former could actually disenfranchise voters who go to the wrong place on election day.

KEEP IN MIND RELEVANCE

As groups work with open data, it may become apparent that some of the data does not offer anything new or informative to their analysis. Much like with all other parts of the observation, analysis, and statement drafting process, groups should be discerning regarding what information and findings they incorporate into public presentations. Organizations should not waste time on information that is likely irrelevant to their observation priorities or may confuse their findings.

STEP 6

MEASURING SUCCESS AND IMPACT

Two data experts look at election data.

It is often easy to get caught up in carrying out your observation activities and not take the time to step back and look at the bigger picture. This is why it is important to plan regular 'reality checks' for program staff to sit down and reflect on successes and shortcomings, make informed decisions, and assess the program's progress in achieving its goals. Reflecting on activities and progress allows groups to assess how well their strategies are working and gives groups valuable information upon which to adjust or improve their data analysis and/or advocacy plan.

HOW EFFECTIVE WAS A GROUP'S STRATEGY TOWARDS ACHIEVING THEIR OBSERVATION GOALS?

For election observer groups, progress will be measured against a group's observation goals and objectives, of which open election data advocacy might be one part of their overall strategy. Under Step 1, groups will have defined observation priorities, objectives, and a strategy that are tailored to their political context. Groups might have different priorities around election data interventions, such as:

- Improving access to election data;
- Improving data literacy among government institutions; or
- Conducting data analysis in order to assess the quality of electoral processes.

Indicators or metrics for measuring progress towards goals and objectives will also look different between groups. Generally, SMART (Specific, Measurable, Attainable, Relevant, and Time-sensitive) goals are better than vague goals, as defining concrete and specific goals will keep a group focused and prevent them from feeling overwhelmed. A group's interventions, including any work around open election data, should contribute to achieving or meeting clear goals.

HOW EFFECTIVE WAS THE ADVOCACY STRATEGY?

Groups that aim to improve access to election data or improve data literacy among government institutions would design an advocacy strategy to advocate for these changes among relevant stakeholders. Under Step 4, groups will have considered approaches to advocating directly with stakeholders to gain access to specific data or data formats. Part of a group's advocacy around election data may include encouraging public institutions to release election data according to open election data principles or improving the way data is communicated to the public. In this case, groups would want to measure the success of their advocacy on increasing access to election data.

Consider any lessons learned: what advocacy strategies worked well? What didn't work well? How will you change the strategies that didn't work well? How else can you adapt? Are you on track to meet your goals?

Potential indicators to measure effectiveness of an advocacy strategy could include: the number of meetings or consultations with government officials and key stakeholders on the importance of open election data, number of resources produced on the importance of open election data, the reach of social media posts/campaigns of the importance of open election data, and any changes in the availability, or format, of official government data. Any indicators used to measure progress should be specific and tailored to the activities and ultimately build up towards the goals and objectives as defined under Step 1.

HOW IMPACTFUL WAS DATA ANALYSIS AND FINDINGS?

Once groups acquire data, groups may analyze this data and use these findings as part of a wider election observation effort to promote electoral integrity. Groups should measure the success of open data analysis against their overall goals and objectives set out under Step 1. Reflection on the success and impact of data analysis will also help groups identify what else can be done with this or other data to bolster traditional election observation methods.

Remember that open data analysis won't always reveal glaring discrepancies or errors in data that should be fixed; finding nothing or very little wrong after analyzing data is a finding on its own.

Consider how well findings from data analysis were used to build accountability or augment other observation analysis. What ways of presenting or otherwise using findings from open data analysis worked well and didn't? Were findings used by other stakeholders, such as the media, government officials, or other civil society organizations?

Potential indicators to measure the impact of open data findings could include: the number of reports or statements issued, the number of media hits or mentions, the number of recommendations made by your group or other CSOs, and any commitments by government officials to address recommendations made by your group.

KEEP TRACK OF PROGRESS OVER TIME

Be aware that what a group may have now is not a guarantee of what they will have in the future. Sustained advocacy is key to creating change around open data. As political leaders, public opinion, and laws and regulations change and evolve, groups will need to adapt to these new challenges and opportunities. As part of this, groups should keep track of progress and impact over time. Measuring success shouldn't happen only once; groups should continuously and regularly assess the success of their advocacy and data analysis effort to ensure they are measuring impact over time, finding and dealing with possibly unintended outcomes and adjusting strategies to achieve better outcomes.

This information can provide useful insights into strategies that can be used in the future for additional advocacy or activities.

REFLECT ON OPERATIONS AND STAFFING CAPACITIES

An important part of evaluating progress is assessing how the group's internal capacity affected its ability to implement its open election data plan. These internal capacities, including staffing and operational restraints, may impact a group's ability to carry out activities and achieve goals.

In Step 3, groups assessed their organizational capacity to use open election data as part of defining an open election data plan. Consider how your internal staffing and operational capacities affected your ability to conduct activities effectively. Are there any improvements in internal expertise or limitations of internal expertise? Did you have the capacities to achieve goals? If not, what could change? What skills, staffing, or other operational considerations are needed to continue progress towards goals? Refer back to Exercise B to think about what may have changed since then.

Use Exercise E: Reflect on Progress as a guide to how groups can structure their reflection on their progress towards established goals.

EXERCISE E: REFLECT ON PROGRESS

Groups should regularly reflect on how their activities are making progress towards achieving their goals. The following questions serve as a guide for a reflection; not all the questions may be relevant to your group's work. Have your open election data monitoring and analysis plan at hand, and start with the category most relevant to your objectives and goals. Make sure to include staff that are directly involved in the management and implementation of your project in your reflection exercise, including those staff working with data directly.

HOW EFFECTIVE WAS THE ADVOCACY STRATEGY?

As part of your broader open election data activities, consider any advocacy your group may have planned to do or embarked on throughout your project to make electoral data more open.

1) Which data was the easiest or quickest to advocate for? Why?

2) What data took more time than anticipated to acquire? Why? What could change in your approach to advocating for this data in the future to get this data faster?

3) Was there any data that your group did not anticipate needing to advocate for but ultimately did?

4) Is there any other data that would be helpful to have in the future for the next steps in your project or for future projects?

5) Which stakeholders, if any, changed in their commitments to open election data?

6) Which stakeholders required or will require more specific or targeted advocacy efforts to achieve your goals? Are there any additional stakeholders your group did not anticipate to engage that should be engaged?

7) What obstacles may influence the availability of open election data in your context?

HOW IMPACTFUL WAS DATA ANALYSIS?

Your group may have gathered and analyzed data throughout your project according to your open election data monitoring and analysis plan. Think about these experiences in the guiding questions below.

1) Did your group experience any unexpected obstacles in retrieving, cleaning, or analyzing data? Are there any skills, resources, or other information your group was missing to successfully collect and analyze data?

2) Did your group notice anything in the data you retrieved that may impact its feasibility, viability, or compliance with open data principles that your group can advocate with stakeholders on?

3) Which findings received the most reaction from civil society or other stakeholders?

4) Which methods of visualizing data or otherwise presenting your findings were most responded to by your audience/s?

5) Which stakeholders committed to or addressed recommendations made by your group based on your findings?

6) To what extent did your open election data analysis compliment your other election-related activities?

7) Are there other ways your findings or data gathered can be used to improve transparency around elections, raise awareness of certain issues, or advocate for specific changes?

ORGANIZATIONAL CAPACITY TO USE OPEN DATA

Looking back at your open election data monitoring and analysis plan and your answers to Exercise B in Step 3, consider how staffing, resources, and internal capacity have changed during your observation effort.

1) Were your resources and internal capacities enough to conduct your activities as planned?

2) Do you see any improvements in internal expertise or capacities?

3) What additional skills, staffing, or other organizational capacities are needed to continue progress towards your goals?

4) How did your group react and adapt to unforeseen challenges or circumstances when advocating for data, finding data, or conducting your analysis? What else is needed to respond more effectively to unforeseen challenges?

Consider the weaknesses or successes in your plan that may be revealed through these guiding questions, and think about whether your open election data monitoring and analysis plan can be changed as needed to improve upon these weaknesses or to continue to use strategies that have worked well for your group.

ACKNOWLEDGEMENTS

This guide was written by Julia Brothers with input from Michelle Brown, Johandra Delgado, Amanda Domingues, and Anastasia Wibawa. Jacqueline Schrag managed the online deployment of the guide, including formatting, graphics and functionality. The design and formatting of this publication was made possible by the United States Agency for International Development (USAID) under Award No. AID-OOA-L-15-00007 through the Consortium for Elections and Political Process Strengthening (CEPPS) in support of emerging groups in the Pacific Islands to use open election data to analyze ongoing electoral processes and promote transparency in the region. NDI would like to acknowledge the work of citizen election monitors around the globe that are embracing open election data as a key component of credible elections, and an important tool for enhancing accountability and public confidence in the democratic process.