Open Data in
Electoral Administration

Open Data in
Electoral Administration

Open Data in

Electoral Administration

Contributors:
Liz Carolan
Peter Wolf

Preface

The concept of open data is based on the idea that data should be freely available for anyone to access, use and share. In recent years, the concept has taken root as an increasingly accepted practice for public and government-produced data, with data seen as a public good, and data infrastructure as the key to accessing it.

Providing election data as open data has numerous benefits: it provides civil society, citizen journalists, electoral observers and citizens access to the same detailed information that was previously only available to selected stakeholders such as large media outlets. In doing so, open data allows all interested stakeholders to follow and understand the electoral process and can enable more inclusive, transparent and trusted elections.

These benefits have been seen in electoral processes from Burkina Faso to the United Kingdom. In Burkina Faso, the real-time publication of election results enabled citizens to compare official results to those of independent observers within hours of polls closing. In Indonesia, efforts led by civil society resulted in the digitization and analysis of thousands of published documents to counter claims of polling fraud. In the United Kingdom, publication of party spending data enabled the identification of mistakes that led to a fine being levied on a major political party.

The open-data concept emerged from the open-source movement, open government and a drive for greater transparency. It has been enabled by wider access to digital technologies for publishing, accessing and processing data, including online. The Internet has enabled the sharing of data at a low cost among a wide variety of people, though equal-access issues remain. In the past five years, many countries have launched open-data initiatives and portals that provide access to various government data sets.

In spite of the potential benefits of open data, election data is often not available in government open-data platforms. If election data is included, it is often limited to election results. Additionally, while most electoral management bodies (EMBs) publish election data directly on their own online platforms, this data often does not comply with open-data principles. Election data may, for example, only be available with restrictive licences, it may be published in formats that cannot easily be analysed or it may not be available at a great level of detail.

In many countries, civil society initiatives, often supported by international assistance providers, have been established to fill this gap. In addition to

analysing and presenting election data, they have also made it available in open formats.

The aim of this publication is to encourage EMBs to increase the application of open-data principles. It provides guidance for improving the practice of data publication towards a more open approach.

Yves Leterme
Secretary-General
International IDEA

Acknowledgements

Acronyms and abbreviations

API	Application programming interface
CKAN	Comprehensive Knowledge Archive Network
CSV	Comma-separated values (file format)
EMB	Electoral management body
FTP	File Transfer Protocol
GEO	Global Electoral Organization
HTML	HyperText Markup Language
ICT	Information and communications technology
ICT4D	Information and communications technology for development
JSON	JavaScript Object Notation (data format)
KPU	Indonesia's Elections General Commission
OD4D	Open Data for Development
ODbL	Open Database License
ODC-BY	Open Data Commons Attribution License
ODI	Open Data Institute
OGL	Open Government License (UK)
OGP	Open Government Partnership
PDDL	Public Domain Dedication License
PDF	Portable Document Format (file format)
RDF	Resource Description Framework
TSV	Tab-separated values (file format)

Contents

CHAPTER 1

Introduction

Open data in short

Open data is data that anyone can access, use and share. It is provided free of charge and free of technical, legal or other restrictions on processing and sharing. Open data is easily accessible, usually online and can be used equally by anyone, including government bodies, civil society organizations, academics, commercial enterprises and individual citizens. Open data should ideally be viewable and processable without needing commercial software packages.

On the other hand, data that can only be accessed by a limited group or individuals, or that is only accessible with restrictions, such as fees, passwords or a requirement to register, is not open data.

Open-data principles

In 2015 the International Open Data Charter convened a global participatory process with civil society, governments, the private sector and multilateral organizations to agree on a set of fundamental norms for public data. They developed six principles that summarize the technical, legal and social aspects of open data. These principles state that open data should be:

1. open by default;
2. timely and comprehensive;
3. accessible and usable;
4. comparable and interoperable;
5. for improved governance and citizen engagement; and
6. for inclusive development and innovation.

EMBs and other governmental bodies embracing open data and the related culture of openness may consider endorsing the Open Data Charter and, in doing so, agree to follow these six principles within their organization.

Background

The concept of open data is rooted in a number of trends across technology and governance. It emerged in Sir Tim Berners-Lee's work to build a linked web of data similar to his original World Wide Web (Berners-Lee 2001; Berners-Lee 2009). This work highlighted the potential of algorithms to pull together data from multiple sources to generate new insights, ideas and services.

The origins of the concept can be found in the open-source software movement, which allows software source codes to be distributed free of charge and encourages researchers to collaborate in order to advance research and development.

The concept's emphasis on transparency led to an integration of its ideas and practices into government commitments to open government and open governance. Here, open data is about improved institutional responsiveness and better decision-making by policymakers and citizens, and it is tied to concepts of citizen choice as a mechanism for improving service delivery (Robinson, Yu, Zeller and Felton 2009; Yu and Robinson 2012).

One result is that a community of civil society organizations has grown around the sector. They discuss the concept as having three main areas of impact: enabling transparency and accountability, increasing the efficiency of public administration and enabling innovation.

These organizations can support institutions wishing to undertake open-data initiatives by providing guidance and training or by providing links with potential local data users, including (but not limited to):

- the Open Data for Development Network (OD4D);
- Open Knowledge International (including its local chapters);
- the Open Data Institute (ODI) (and its network of local nodes);
- GovLab, which maintains a database of impact case studies; and
- sector-based projects and coalitions, such as the Open Contracting Partnership and the Global Open Data for Agriculture and Nutrition Initiative.

Open government data

As part of their work, government institutions compile and produce a great amount of data. This data can be of great value for users outside government,

not only for developing a better understanding of how the government operates, but also for encouraging more public participation and for influencing and improving decision-making. Governments also hold key data assets that can enable individuals, start-ups, academics and companies to develop services, knowledge and insights, ranging from geo-spatial, satellite and infrastructure data to demographics and economic data.

Open-government principles have been increasingly adopted around the world, thanks in large part to the advocacy efforts of civil society organizations at national and global levels across the transparency, governance, civic-tech and technology-for-development (ICT4D) fields.

In 2011 eight governments (Brazil, Indonesia, Mexico, Norway, the Philippines, South Africa, the United Kingdom and the United States) founded the Open Government Partnership (OGP) and endorsed the Open Government Declaration (Open Government Partnership 2011). Since then, 66 more countries have endorsed the declaration, taking the total number of participating states to 74. Many of the OGP countries have explicitly committed to open data, often starting by creating the necessary technical building blocks, including open-data platforms, machine-readable formats, licences and permissions, and, in a few cases, focusing on what data should be published and how to make sure the data is actually used (Foti 2015).

As part of their open-government initiative, many countries have established technical open-data infrastructure and expertise at the national and sometimes also the regional or local level that may also be utilized for electoral purposes. As of 2016 a comprehensive list of such open-data infrastructure contained over 500 open-data portals (Open Knowledge International (a)).

Characteristics of open data

Data is open if it can be freely accessed, used, modified and shared by anyone for any purpose, subject only, at most, to requirements to provide attribution and/or to permit distribution under identical open conditions. Specifically, open data is defined by the principles set out in the Open Definition (Open Knowledge International (c)) and requires that data be:

- legally open, i.e. available under an open (data) licence that permits anyone to freely access, reuse and redistribute it; and
- technically open, i.e. that data be available for no more than the cost of reproduction and in machine-readable and bulk form.

According to Open Knowledge International, the key features of openness are:

- Availability and access: data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading

it over the Internet. Data must also be available in a convenient and modifiable form.

- Reuse and redistribution: data must be provided under terms that permit reuse and redistribution, including the intermixing with other datasets. Data must be machine-readable.
- Universal participation: everyone must be able to use, reuse and redistribute data—there should be no discrimination against fields of endeavour or against persons or groups. For example, non-commercial restrictions that would prevent commercial use or restrictions of use for certain purposes (e.g. only in education) are not allowed (Open Knowledge International (i)).

The format in which data is published also greatly affects its openness. Data published in formats that are not machine-readable, e.g. Portable Document Format (PDF), are considered the weakest form of open data, according to Berners-Lee's ranking of the 'five stars of open data' (5 Star Open Data 2012), and should only be used where other formats are not feasible. Proprietary formats, such as Excel spreadsheets, are better, as they allow for processing, but non-proprietary formats, such as comma-separate values (CSV), are preferable, as they do not require licensed software to process. The top of Berners-Lee's hierarchy is reserved for formats such as Resource Description Framework (RDF) that enable linked open data (see World Wide Web Consortium n.d. for details).

Some definitions of open data, including those in some government policies (Larrick 2016), consider only non-textual material, such as databases, statistics and maps that are made openly available, to be open data. However, textual data is quite often as important to data users, and should be published openly in the formats most appropriate to enable its use. In the context of elections, this includes texts such as electoral legislation, EMB meeting minutes and decisions, as well as voter information material.

From a technical point of view, the Open Government Working Group (2007) identified the following as requirements for open data:

1. **Complete**: All public data is available that is not subject to valid privacy, security or privilege limitations, as governed by other statutes.
2. **Primary:** Data is published as collected at the source, with the finest possible level of granularity, not in aggregate or modified forms.
3. **Timely**: Data is made available as quickly as necessary to preserve its value.
4. **Accessible:** Data is available to the widest range of users for the widest range of purposes.

5. **Machine-processable**: Data is reasonably structured to allow automated processing of it.

6. **Non-discriminatory:** Data is available to anyone with no registration requirement.

7. **Non-proprietary:** Data is available in a format over which no entity has exclusive control.

8. **Licence-free:** Data is not subject to any copyright, patent, trademark or trade-secret regulation. Reasonable privacy, security and privilege restrictions may be allowed as governed by other statutes.

Later, the Sunlight Foundation (2010) published an expanded list with ten principles by adding:

9. **Permanence**: Once data is available online, it should remain online, with appropriate version tracking and archiving.

10. **Usage costs:** No usage costs should be imposed on the public for access.

Finally, publishing data in accordance with recognized international standards allows for greater interoperability and the combination of data sets from different sources. The Open Data Institute (ODI) published a conceptual model informing the design of open data standards for election (Dodds 2014).

The data spectrum: open, closed and shared data

As highlighted above, not all data is suitable and available for complete and open publication. Privacy considerations and legal limitations may, for example, be factors that limit the openness of a given data set.

To provide a framework for discussing the openness of data, the ODI introduced the data spectrum (Open Data Institute n.d. (b)), according to which the openness of data can be classified as:

- **closed, internal access only:** for internal use of the institution holding the data;
- **shared with named access:** available for explicitly named external data users;
- **shared group-based access:** available to a greater, but restricted, audience; restrictions are implemented via authentication mechanisms;
- **shared with public access:** anybody can access the data, but licence restrictions limit the permitted use of the data; or
- **open for anyone:** no legal or technical limitations for anybody to use the data.
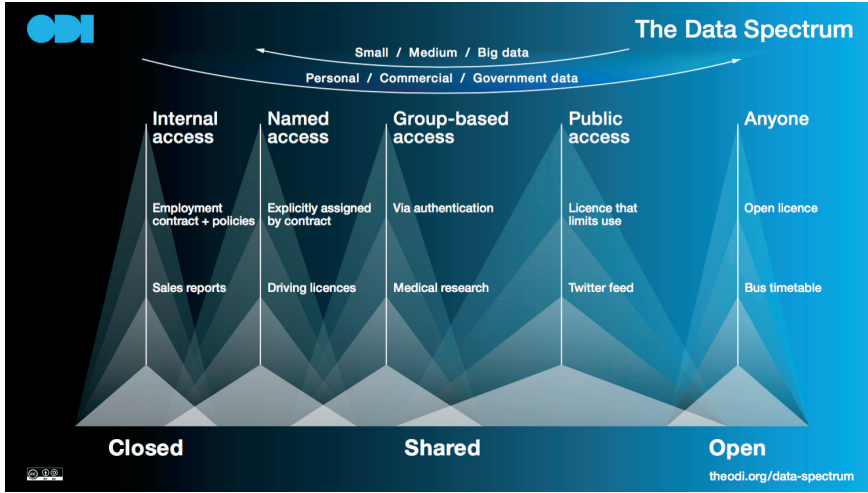
Figure 1. The Open Data Spectrum

CHAPTER 2

Open data in elections

Open data in elections

Democratic processes such as elections and referendums are successful when the electorate can make informed decisions; when trust in the process is maintained; when fraud is disincentivized, detected and punished; when uncertainty surrounding results is minimized and when the system is enabled to continually improve. Open data can contribute to making this a reality by enabling greater transparency and by helping build data infrastructure that enables the multiple actors involved in elections to carry out their work.

Transparency is a key requirement for genuine, competitive and democratic elections. It is the basis for public confidence in the electoral process. **The ultimate goal of open data in elections is to increase electoral integrity and accountability through more transparent elections**. Open data improves transparency by providing all citizens unrestricted access to electoral data, by allowing for timely and detailed analyses and by not discriminating against any individuals or groups (Open Knowledge International n.d. (j)).

Burkina Faso held its first election since the ousting of the country's long-time leader in November 2015. Through a carefully constructed results transmission system and web application, the Election Commission worked with the national open-data team to release real-time, verified results district by district as an open data set and on an online portal at <http://www.burkina2015.bf>. Media sources that were briefed in advance picked up the data and distributed it to citizens and interested parties inside and outside the country. An independent parallel vote count was conducted by Codel, a coalition of civil society organizations, which enabled early results to be quickly checked and verified. The winner of the election was announced a little over 24 hours after the polls closed, and the result was accepted by the losing candidates (Scott 2016).

For data providers, including EMBs and other institutions in charge of election-related information, applying open-data principles provides an **opportunity to demonstrate their commitment to transparency and accountability**.

Open data also delivers information symmetry and thus gives **more equal power to electoral stakeholders**. Multiple actors are needed for elections to function, including EMBs; political parties, candidates and campaign teams; civil society; electoral observers; journalists; analysts, academics and interest groups.

Each of these actors needs data to be able to perform their role properly. We can therefore think of electoral processes as needing data infrastructure: core pieces of data that key actors should be able to access and use to do their jobs.

For these **data users**, open data assures the **best possible and most efficient access to the data** needed for their analysis. Among other things, open election data can:

- inform voters about when, where and how they can vote;
- enable electoral observers to compare official election results with their observations at polling stations;
- enable interested stakeholders to tabulate election results and compare them with the official tabulation;
- allow interested stakeholders to follow the tabulation process in close to real time;
- enable interested stakeholders to conduct voter list audits and to assess the quality of voter lists;
- enable civil society organizations to better plan strategies to increase the amount of information available to citizens;

- show electoral stakeholders how electoral boundaries are drawn and assess the rationale behind this; and
- provide access to complex data sets, such as political finance data, for efficient analysis.

Finally, there are already **many good examples of the successful use of open elections data**, from Afghanistan (Afghanistan Election Data n.d.) to Burkina Faso (Scott 2016), and from Indonesia (Graft Verhulst and Young 2016) to Tunisia (Schuller 2014) to the United Kingdom (Electoral Commission 2016(a)). One of the lessons learned from these cases is that opening up data so any actor can access, use and share it can, in some instances, improve the functioning of elections.

Case Study: Campaign Spending Data in the United Kingdom

In January 2016 the UK Electoral Commission published the Labour Party's 2015 general election spending return on its online registers database. Immediately following publication, the Commission received enquiries from journalists who had analysed the data and identified something the Commission's own analysis had missed. The journalists were unable to find any reference to spending incurred on a stone tablet used to promote the party's pledges during the campaign, which had been the subject of some media attention. Based on this outside analysis, the Commission launched an investigation two days later, which then uncovered some further (unintentional) mistakes made by the party. The Commission found that the party's campaign spending return did not contain all campaign payments, as is required by UK law, and imposed a financial penalty of GBP 20,000 (Electoral Commission 2016(b)).

Which data to publish

Applying open-data principles in electoral processes requires an understanding of which electoral data sets should be openly available.

The Global Open Data Index (Open Data International n.d.(e)), an annual survey to measure the state of open government data around the world, tracks only one type of electoral data, i.e. election results. While results data is clearly a first indicator of the openness of election data, there is far more information and data that can be made available about the electoral process.

A very comprehensive definition of data sets that can be considered for publication as open election data has been established by the Open Election Data Initiative (National Democratic Institute 2015(a)). The Open Election Data Initiative's definition includes tabular and spatial data, as well as textual data. With this definition, open-data principles can be applied in all phases of the electoral cycle. The data sets covered are mostly those for which EMBs are responsible, but they also go beyond that:

1. **boundary delimitation:** granting access to population data, as well as the spatial data for electoral districts derived from population data (data tables and maps);

2. **political party and candidate registration:** granting access to lists of parties and candidates that applied, were registered and were rejected by the EMB, also including registration information;

3. **campaign finance:** information about the funds parties and candidates received and spent for their campaigns, including limits and regulations;

4. **voter registration:** statistics about registered voters, including breakdowns by gender, age and geography, also including information about the registration process itself;

5. **voter lists:** access to detailed information about eligible voters, including names, dates of birth, national ID number, residence, polling station, etc.;

6. **polling stations:** including addresses, contact information, number of voters for each polling station;

7. **election results:** including the number of registrants, valid and invalid votes, as well as votes for each party and candidate. All this data should be available disaggregated down to the lowest level at which votes are counted;

8. **legal framework:** granting access to all laws, regulations and instructions related to the electoral process (textual data);

9. **electoral management body:** access to contact information and details for officials at all levels (textual data);

10. **EMB processes:** data about EMB decisions, minutes and resolutions;

11. **election campaigns:** data about timetables for campaigns, related regulations and restrictions, as well as public resources for campaigns;

12. **voter education:** access to information for voters about parties, candidates, registration and voting procedures;

13. **electronic voting and counting:** provision of details regarding procurement, accuracy, security and reliability of the voting system, also including source codes;

14. **electoral complaints and dispute resolution:** information about the number and types of complaints, the parties filing them, the outcomes of dispute resolution, as well as details about accessing the complaints process;

15. **election security:** information from regulatory, security and judicial institutions about standards for police and military in the electoral process.

Principles to guide the release of open election data

Developing a plan for releasing open election data requires balancing multiple factors, such as the need for timeliness with the need for accuracy or the need for granularity with the need to protect privacy. This may require analysis and deliberation regarding the different types of data concerned.

The Open Election Data Initiative established nine principles (National Democratic Institute 2015(c)) that can guide the release of open election data and support this deliberation process. According to these nine principles, open election data should be:

1. **timely**: made available as quickly as necessary to be useful;
2. **granular**: available at the finest-possible level or granularity or detail and also made available at the primary level, i.e. the level at which the source data is collected;
3. **available for free on the Internet:** released without any monetary restrictions;
4. **complete and in bulk:** released as a complete data set without any omissions;
5. **analysable**: made available in a machine-readable format that can be quickly and easily analysed;
6. **non-proprietary:** available in a format over which no entity has exclusive control;
7. **non-discriminatory:** available to any individual or institution for anonymous access without any usage restrictions, including application or registration requirements;
8. **licence-free:** there should be no barriers for reuse and redistribution for any purpose;
9. **permanently available:** via a stable Internet location for an indefinite period of time. Specifically, data that is only available for a short period of time is not open.

CHAPTER 3

Current practice for the publication of election data

Even though disclosing election data has many potential benefits, relatively little election data is published according to open-data principles. From available sources, it appears that, in spite of an increased discussion about open election data, only limited progress has been made in this field in recent years.

Availability varies depending on the type of election data in question. Registration- and results-related data is more commonly publicly available than political finance information, for example.

Even where election data is available, the openness of this data is in most cases restricted to an extent in that it does not comply with open-data principles. On the ODI data spectrum, election data in the 'open' category is more an exception than the rule. Where election data is not 'closed', it is commonly in the 'shared' category, where limitations include time and duration of data availability, the granularity of available data, restrictive licensing, non-machine-readable data formats, and registration and authentication requirements for data users.

The Global Open Data Index

Since 2013 the Global Open Data Index (Open Knowledge International n.d. (f)) has tracked and scored the openness of countries' data, including data concerning election results. The Index only focuses on results data, which is arguably the most essential election data to be provided openly. Of the 122 countries it now covers, only 14 received a score of 100 per cent in 2015, 15 received a 100 per cent score in 2014 and 13 received a 100 per cent score in 2013.

This stagnant trend does not change when looking at countries with a score of 70 per cent and above: 36 countries received this ranking in 2015, 35 countries in 2014 and 29 countries in 2013.

The Index is crowd-sourced, so it may not necessarily be a completely accurate reflection of each country's status. However, it indicates that the number of countries employing good open-data practices in elections is still very limited and that not much progress was made over the years covered.

Open Election Data's Latin America Data Inventory

The Open Election Data Initiative has developed the Latin America Data Inventory (National Democratic Institute 2015(b)), a database of the openness of election data across 18 countries in the region as of 2014. The section about results data renders a similar picture as the Open Data Index: only one of the 18 surveyed countries was assessed 'mostly open', eight as 'partially open' and nine as 'not open'.

The 13 other categories the Inventory covers show a mixed picture for open election data in the region. In more than half of the surveyed Latin American countries, voter education material, voter registration lookup, contact data for EMBs and polling stations and data about complaints and appeals were openly available.

However, party and candidate registration data, especially about rejections, campaign finance data and boundary delimitation details were less commonly published in open-data formats.

International IDEA's ICTs in Elections Database

International IDEA's ICTs in Elections Database includes global information about online publication practices by EMBs based on a global survey of over 170 countries.

According to the database, the individual verification of voter registration status through online lookup tools is very common: 52 per cent of all countries provide such data access to citizens. Such systems are very common in the Americas (79 per cent) and Asia (62 per cent), but less common in Europe (46 per cent) and Africa (39 per cent).

Interestingly, privacy concerns do not prevent the online publication of full voter lists, including personal details of the entire electorate in 14 countries.

Does the country provide individual online voter registration/polling assignment checks?

ICTs in Elections Database - Online data publication by EMBs

- Yes, confirmation of registration using online interface
- Yes, assigned polling station using online interface
- Yes, confirmation of registration using mobile phone
- Yes, assigned polling station using mobile phone
- No
- Not applicable
- Multiple answers
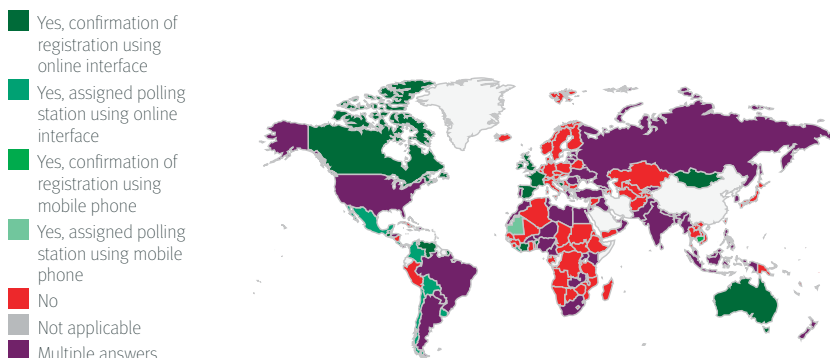
Figure 2. Provision of online voter registration lookup tools for individuals
Source: International IDEA <http://www.idea.int/data-tools/question-view/754>

Voter registration statistics are openly accessible in 66 per cent of all countries worldwide and are especially widespread in Europe, where such data is available online in every country.

Does the country publish the statistical overview of voter lists online?

- No
- Yes, at polling station level
- Yes, at the municipal level
- Yes, at the constituency level
- Yes, at the national level
- Not applicable
- Not specified
- Multiple choices
- Other
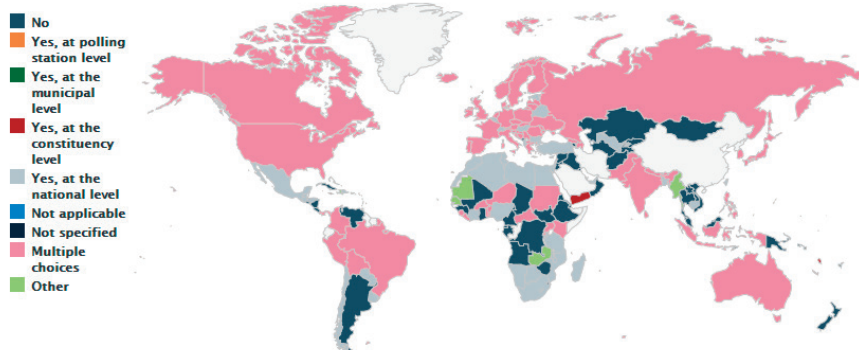
Figure 3. Online publication of voter registration statistic
Source: International IDEA <http://www.idea.int/data-tools/question-view/755>

Election results are published online in 81 per cent of the surveyed countries; however, the openness of this data varies greatly, and only a few cases fulfil all open-data principles:

- only 38 per cent of countries publish data in a machine-readable format;
- only 26 per cent of countries publish data at the polling station level;
- only 20 per cent of countries publish data as results become available to the EMB; and
- data about historic results is accessible in 57 per cent of all cases.

Party and candidate financing reports are published in only 32 per cent of the surveyed countries and in only 6 per cent of all cases in machine-readable formats.
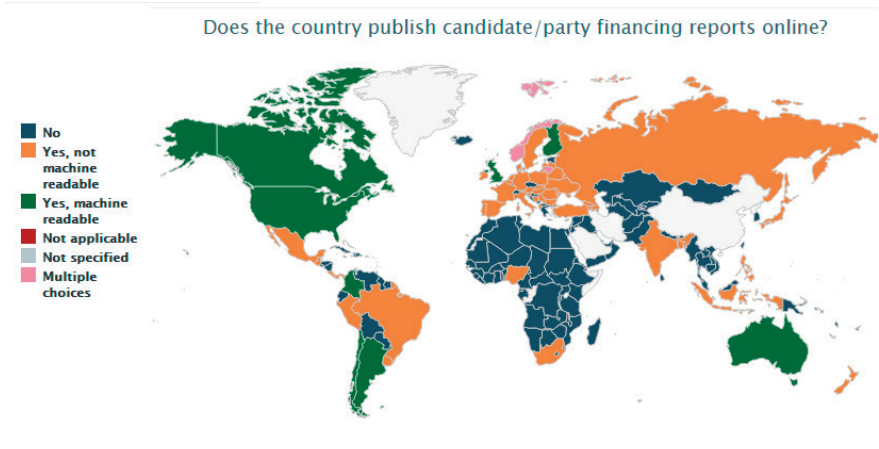
Figure 4. Online publication of candidate and party financing reports
Source: International IDEA <http://www.idea.int/data-tools/question-view/765>

CHAPTER 4

How to initiate an open-data initiative in the field of elections

With much election data still not openly available, EMBs can and should do more to open the data sets they maintain. Disclosing election-related data, as in other areas of government, is in part a technical process, but it also requires a cultural and institutional change.

Assessing institutional readiness and progress

Any organization, such as an EMB, wishing to start an open-data initiative, should first identify if there are already existing open-data initiatives and projects being implemented in their country. A good starting point is to determine whether the country is a member of the Open Government Partnership and, if so, identify who the point of contact is according to the OGP website (see Open Government Partnership 2016). An alternative is to determine if there is a national data portal and, if so, who manages it, or if there is a local civil society chapter of the Open Knowledge International in operation (see Open Knowledge International n.d. (h)).

There are a number of tools available to assess an institution's readiness for, and progress with, an open-data initiative. The tools that are currently available are generic and not specific for EMBs and election-focused organizations; however, they can provide an understanding of the starting point and of actions to prioritize.

The first is the Open Data Readiness Assessment, a World Bank tool that can be used to conduct an action-oriented assessment of the readiness of a government or individual agency to evaluate, design and implement an open-data initiative (World Bank n.d.). Freely available in multiple languages, this tool is intended for governments starting afresh in the field of open data, and usually requires several days of an analyst's time to conduct interviews and

complete an assessment. It poses questions on an institution's preparedness under the following headings:

1. Legal and leadership framework;
2. Institutional framework;
3. Data within government;
4. Demand for open data;
5. Citizen engagement; and
6. Infrastructure (Internet and ICT).

The Open Data Pathway (see Open Data Institute n.d (a)) is a self-assessment tool that enables an organization to assess how well it publishes and consumes open data, and to identify actions for improvement. The tool uses a questionnaire based on the open-data maturity model (see Open Data Institute 2015(b)). This is intended for use by individual organizations for monitoring progress towards open-data implementation. It can be completed by someone with knowledge of the relevant organization's data policies, procedures and skills in a single sitting. The model is based around five themes and identifies progress on five levels. The five themes are:

1. data management processes;
2. knowledge and skills;
3. customer support and engagement;
4. investment and financial performance; and
5. strategic oversight.

Cultural and institutional change

The energy around an open-data project can sometimes focus solely on technical implementation. However, in order for an initiative to be successful, cultural and institutional change is required.

The Open Data Institute published a guide in 2015 called 'Open data in government: how to bring about change' (see Open Data Institute 2015(a)). The guide suggests that the challenges of implementing an open-data initiative are similar to any organizational-change programme, and it highlights the need to build support and consensus for the initiative and to identify ways to overcome resistance to change. The 12 recommendations they make for sustaining a change to open data are:

1. Articulate your vision, with clear examples of the benefits open data will bring.
2. Secure support for your open-data initiative from both senior/political leadership and officials within departments before launch.

3. Combine top-down leadership for your open-data initiative with support for individual or front-line champions.

4. Build open communication and mechanisms for feedback into your open-data initiative from the outset, both inside and outside your organization.

5. Set out some quick wins for your open-data initiative, such as supporting a pilot test case.

6. Be flexible and responsive to the strengths and needs of different teams involved.

7. Consolidate your change-management efforts.

8. Ensure that there are people responsible for change management as part of your team.

9. Seek and foster stories of the impact of open data.

10. Foster external support within civil society and academia to drive continued demand for data.

11. Introduce opportunities for officials to take part in ongoing learning about open data.

12. Build metrics to regularly evaluate your open-data initiative.

One step commonly taken by institutions aiming to embed a change in the direction of greater publication and use of open data is to draft and release an organizational policy. ODI guidance highlights the four elements of a good open-data policy as:

1. a definition of open data, why it is important for the organization, and the reasons for the policy;

2. a general declaration of principles that should guide the release and reuse of open data;

3. an outline of the types of data collected by the organization and whether they are covered by the policy; and

4. references to any relevant legislation, policies or other guidance that also apply to the management and sharing of information with third parties (Open Data Institute 2016).

Creating a culture of openness

Disclosing election data requires the development of a culture of openness within the institution providing the data. This culture entails a commitment to transparency, combined with recognition of the potential for data misuse and of the risks related to disclosing data.

Open data changes an EMB from a data owner into a data provider. As a data provider, an institution not only becomes more transparent, it also opens up new avenues for feedback and criticism that need to be addressed accordingly.

Mistakes happen in any institution. With higher levels of transparency, they get exposed more easily and more quickly. When publishing large and detailed data sets, especially in real time and with necessarily limited quality checks, there is always a possibility that some mistakes will remain undetected before publication. Even a few exposed mistakes can create the impression that there are more, still undetected, problems in the electoral process. Some actors can also be expected to look for such mistakes in an attempt to undermine or discredit the organization.

If such situations are not handled well, the result can be the opposite of what was intended: instead of increasing integrity, distrust in institutions and processes may be increased. This needs to be addressed in several ways: the EMB needs to proactively disclose and communicate any limitations of the open data they provide and misunderstandings that may arise, and the ethical use of the data provided needs to be promoted. Where data is preliminary or partial, this must be clearly explained. Most importantly, the EMB must be very responsive to complaints, requests for corrections and feedback from the public, political parties and others.

Open data and elections: sensitivities and risks

The potential of open data for electoral processes must be balanced with an understanding of some of the sensitivities and risks related to data release.

Balancing transparency and privacy

As with most sectors, some data pertaining to elections may be sensitive and may need to remain closed, to be shared with only a few parties or to be disclosed only above a certain level of aggregation. Voter registers, for example, contain private information that needs to be managed properly. The data spectrum, which sorts data on a spectrum from 'closed' to 'open (outlined above), provides a helpful conceptual framework for this thinking.

These risks are especially important in the context of vulnerable populations. In a recent case in India, the government was forced to remove private information—including their address, age and contact number—about single women taking part in government welfare programmes that was inadvertently made public, due to concerns that this could lead to harassment (Gowda 2015).

The OGP-produced 'Open Government Guide' places the responsibility on institutions to be accountable in their handling of citizens' personal information. They state that the two information rights—the right to privacy and the right to information—must be balanced and work in tandem if the impacts from greater transparency are to be meaningful (Open Government Partnership n.d.).

At present, however, access to election data is not organized well in developed or developing countries, as seen from recent examples from Afghanistan (Development Seed 2014), the United Kingdom (Doyle 2014) and the United States (Kaplan 2016) . Even when managed well, the ability to access data can be based on the ability to pay or other criteria that enhance incumbency advantage or exclude smaller, less-resourced parties from analysing it.

Risk of the misuse of data

A concern that is sometimes raised by those considering releasing data is that it might be misused, misinterpreted or misrepresented.

There are fears that data published will be used by those wishing to carry out unlawful or disruptive activities. For example, one African country withdrew water infrastructure data from its portal due to fears it could be used in order to contaminate the water supply, and that responsibility would land on the desk of the institution that published the data.

A second fear is misinterpretation of data. In electoral processes, this can be seen when early results, released in real time, can give a misleading indication of who will win overall. In Ethiopia in 2005, for example, early results, which came from the opposition-friendly capital, were interpreted as indicating a landslide victory for the main opposition party. When the final results showed a different outcome, this may have contributed to mistrust in the process, mistrust that impacted the period of uncertainty following that election (International Foundation for Electoral Systems 2005).

A further concern occasionally voiced by data holders is the misrepresentation of data, such that it is either taken out of context or falsely attributed. This can include misleading news items critical of institutions or the misattribution of falsehoods, such as statistics, to an agency or organization.

There is little that can be done from a technical point of view or in the way that data is published to address these concerns. These are risks that need to be addressed within the broader policy and legal framework of a polity. Officials who release data should be given the cover and security of a solid policy that backs up their decisions to publish data. Election organizers should ensure that every effort is taken to provide accurate information to counter misinterpretation, and should be supported by legislation or guidance that allows sanctions against those misleading the public.

These risks can be mitigated somewhat through careful consideration and planning behind decisions to publish data, which should include consultations with those wishing to use the data and those with an understanding of the sensitivities and risks involved.

Overall, however, it should be acknowledged that an open-data initiative will not, on its own, provide a 'magic bullet' answer to improving the functioning of electoral processes, in particular making them more transparent and trusted. Such an initiative can be a vital, but insufficient, factor, which is most impactful when combined with other governance-building initiatives (see Carolan 2016(a)).

Stakeholder involvement

Open data is ultimately about enabling improvements in information flows; the creation of new knowledge, insights and services; and greater transparency of processes. For this to be realized, the data has to be picked up and used by actors beyond the data-holding institution.

Case Study: Mapping Election Data in Afghanistan

In the weeks after the 2009 election in Afghanistan, a group of technologists built a map using released open election data that was able to quantify the extent of fraud. It showed that a single candidate received over 90 per cent of the votes in many polling stations that returned a high ballot count. The map ultimately informed the decision to force a run-off. The project was repeated in 2014, and was expanded to include a data exploration tool that made it possible to identify anomalies. In one region, results showed candidate Ashraf Ghani winning 100 per cent of 600 ballots cast at well over 100 polling stations (Development Seed n.d.).

The surest way to enable stakeholders to access and use data is to involve potential users in the planning and design of the initiative, including the prioritization and periodization of data sets, and the selection of formats and data publication protocols. It is most often these stakeholders who will turn data into information, and they may not do so if the data selected for publication is not useful to them, or if they lack the required skills to process the data.

Important electoral stakeholders include the media, political parties, civil society, academia, electoral observers, EMBs and possibly other governmental bodies in charge of maintaining and processing election data. They all play different roles in an open-data environment:

- **Data providers** are the institutions or organizations that maintain and release data sets. In electoral processes, this may often be the EMB or other government bodies in charge of electoral activities such as voter registration, boundary delimitation or electoral justice. It can also include other stakeholders such as political parties, electoral observers or researchers and analysts.

- **Data consumers** are those who directly utilize the data provided, as open data is accessible to anybody without restriction; every citizen is a potential open-data user.
- **Data intermediaries** often sit between data providers and the public. While open election data is in principle available to anybody, using open data sources requires skills, and especially large data sets may be too difficult for direct use by the average citizen. When data processing, analysis and presentation skills are required, efficient use of open data commonly requires intermediaries (Whong 2015). These may be data journalists, data scientists or civic coders who turn the data into insights, including visualizations, maps or news stories, reports and analysis. In electoral processes, data intermediaries often include civil society, media and electoral observers.

Sometimes stakeholder roles can be blurred:

- **Data intermediaries opening up datasets in closed formats**: when original election data is not provided in open formats, intermediaries may, in addition to analysis, also take on the task of collecting and converting originally closed data sources such as PDF documents into open data. Examples of this include the Tunisia Election Data project (Mourakiboun n.d.) or Kawal Pemilu in Indonesia (ACE Project n.d., see box below). This additional step and task can be avoided by the original data being provided already in open formats, allowing intermediaries to focus solely on analysis and visualization.
- **Data providers also providing analysis tools**: data providers, such as EMBs, may also take on the additional task of providing user-friendly analysis tools on their open-data portals. This can be done to provide services directly to citizens, to compensate for a lack of data intermediaries, or in collaboration with intermediaries, as was the case with the open-elections project in Burkina Faso (Scott 2016). Providing tools alongside data in the early stages of an open-data initiative is an increasingly common practice for portals (see, for example, the Tanzania Government Open Data Portal n.d.) in order to demonstrate the potential uses of the data and to build momentum for the initiative.

Indonesia's General Elections Commission (KPU) announced it would begin sharing all election-related data on the Internet in April 2014 following calls from citizens and political parties to improve transparency and safeguard that year's presidential election.

The results of that election proved to be contentious, and the two contenders traded allegations of vote rigging. The KPU published the voter tabulation forms from various levels of the voting process, but only made the polling place tabulation available as image files, so they could not be processed. A group of technologists and volunteers created a project called Kawal Pemilu (Guard the Election) with a website that allowed citizens to compare official vote tallies with the original tabulations from polling stations. The tabulations were made public, but 700 volunteers had to digitize the forms and make the data more accessible.

GovLab's case study found the impact of this data sharing to include enabling citizen participation, increasing trust in official tallies and easing the democratic transition (Graft, Verhulst and Young 2016).

When engaging with electoral stakeholders, the following questions can help guide this approach:

- Who is interested in using open election data? Is there already an audience demanding access to open election data or does this audience need to be built first?
- What data sets are potential users most interested in? How should the publication of different data sets be prioritized?
- Do potential users have the necessary technical skills to use open data sources?
- Do potential users have enough electoral expertise to understand the provided data?
- Is there a risk of misunderstandings and misinterpretation among data intermediaries and consumers?
- Which explanations are essential for users to fully understand the data?
- What are the most suitable data formats for stakeholders and the data to be published?
- Do potential users require additional training?
- Can such training be provided by the EMB or other organizations?

Getting open-data expertise

When an EMB embarks on an open-data project, it is unlikely to be the only institution in the country that does so. It is therefore useful to map domestic open-data expertise and infrastructure, and, when necessary, to also consider

getting support from international experts and assistance providers. Open-data initiatives to reach out to may include:

- the Open Election Data Initiative, which provides guidance specific to electoral processes, administration and data;
- Open Knowledge International, which has a global network of locally established affiliate organizations that often include civil-tech projects and expertise;
- the Open Data Charter, a coalition of open-data organizations and governments that build partnerships with sector leaders wishing to develop protocols, standards and pilot projects for open data in different domains;
- the Open Data Institute (and its network of local nodes), which has developed guidance and tools for institutions and can provide technical support and assistance.

CHAPTER 5

Technical aspects of implementing an open-data initiative

The 'Open Data Handbook' (Open Knowledge International n.d. (b)) provides very detailed technical guidance on how to open up data in general. While not all recommendations apply equally to election data, they are a good starting point, and this chapter uses them as a basis for discussing the specific needs for making election data open.

The 'Open Data Handbook' recommends the following key rules (Open Knowledge International n.d. (g)):

- **Keep it simple.** Start out small, simple and fast. There is no requirement that every data set be disclosed right now. It is fine to start out by opening up just one data set or even one part of a large data set. For EMBs, starting small in this way will be most effective if data is released along with some planning around how it might be used, such as a visualization or partnership with a potential user (see the example of Burkina Faso (Scott 2016)).
- **Engage early and engage often**. Engage with actual and potential users and reusers of the data as early and as often as you can, be they citizens, businesses or developers. This will ensure that the next iteration of your service is as relevant as it can be. Much of the data will not reach end users directly but rather via infomediaries, i.e. people who take the data and transform or remix it for presentation.
- **Address common fears and misunderstandings**. This is especially important if you are working with or within large institutions such as government. When opening up data, you will encounter plenty of questions and fears. It is important to (a) identify the most important ones and (b) address them at as early a stage as possible.

When applying these rules to election data, EMBs are likely to face specific challenges connected to the fast-paced nature of electoral processes. Interest in, and requests for, election data can be expected to be highest in the electoral period, a time when there is not much space for making mistakes and when correcting mistakes in already published data can cause controversy. The recommendation for effective stakeholder engagement and addressing possible misunderstandings is arguably even more essential with election data than it is for other data sets.

In addition to the key rules, the 'Open Data Handbook' lists the four basic steps for making data open:

1. Choose the data set(s) you plan to disclose. Keep in mind that you can (and may need to) return to this step if you encounter problems at a later stage.
2. Apply an open licence.
    a. Determine what intellectual property rights exist to the data.
    b. Apply a suitable open licence that licenses all of these rights and supports the definition of openness discussed in the section above called 'What is open data?'
    c. NB: if you are unable to do this, go back to step 1 and try a different data set.
3. Make the data available in bulk and in a useful format. You may also wish to consider alternative ways of making it available such as via an application programming interface (API).
4. Make it discoverable: post it on the web and perhaps organize a central catalogue to list your open data sets.

Choosing data sets

As a first step, it is necessary to create an inventory of all the data an EMB has available and that may be suitable for open publication. The categories outlined in the section 'Open election data' may serve as a starting point for this. Some of the identified data sets are likely to have already been published online, but they may not adhere to all open-data principles.

As most EMBs will already have some election data publicly available, the data inventory can include checking, for each data set, which open-data principles are not yet being adhered to. This entails investigating common limitations, including whether the data is openly licensed, machine-readable, publicly available in bulk, timely and up to date. Where these limitations are in place, the necessary steps to change this should be analysed and assessed for feasibility.

For each data set, it is also necessary to determine the formats in which it is or will be available internally. Is the data available electronically? Is the data available at the granular level? Detailed polling station results are frequently only available at the local level in paper form, and the central election administration has aggregated results only at a higher level.

Additional questions include the following. Is real-time provision of the data possible? Can data be published quickly, possibly before any quality checks, without releasing inaccuracies that may undermine the credibility of the election? Is data available in machine-readable formats? If this is not the case, is it feasible to convert the data into formats suitable for open publication?

Gender and diversity data should also be reflected in the data inventory. Wherever gender-disaggregated data can be released, recognizing and addressing patterns of marginalization will be facilitated (Bardall 2015).

Restrictions and limitations

With the data inventory available, the next issue is to determine which data is suitable for what type of publication, i.e. which data should be open, which data can only be shared with selected stakeholders and which data needs to remain internal. While restrictions should not be conceived as an excuse to avoid the open publication of data, there are some valid reasons why not all data can be fully published:

- There may be **legal restrictions that prevent the publication of some detailed election-related data** if, for example, the EMB does not own all the data it is using, if this data is sourced or licensed from other institutions or companies.
- **Privacy issues always need to be considered when publishing open data**. This is especially true for voter registers that contain personal information for large parts of a country's population that should not be published without restrictions. While some countries publish voter registers in full detail (International IDEA n.d.), most apply certain restrictions. Restrictions include releasing voter registers to a limited set of stakeholders, publishing in formats that are not easy to analyse, publishing only abridged versions of the register, excluding details such as dates of birth, ID numbers, addresses, etc.
- The publication of **election results from small voting districts or segments of voters** that allow conclusions about voting patterns of individuals or communities can endanger the secrecy of the vote and create a danger of retribution against (or compensation for) such communities for obtaining a certain election result.
- Other reasons for not publishing data openly may include threats to economic interests or the **security of the country or electoral process**.

Setting priorities

The data inventory may indicate many issues that require significant and costly changes to data transmission and processing before election data can be published according to all open-data principles. It is unlikely that all available data will be openly published immediately, so priorities need to be established at this stage.

Establishing priorities is about balancing a number of factors, two of which are discussed above: the specific context, including any **risks or sensitivities** pertaining to the electoral process and its data; and **stakeholder needs**, in particular those of potential data users. A third factor, as outlined in the 'Open Data Handbook', is the '**ease of release**'.

The 'Handbook' says that rather than deciding which data would be most valuable, it can occasionally be useful to take a look at which data is easiest to get into the public's hands. Small, easy releases can act as a catalyst for larger behavioural change within organizations. However, a counterpoint is that small releases that are of such limited value that nothing can be built from them can undermine faith in the entire project.

When prioritizing between the various open-data principles, key principles to focus on first may include timeliness, free access and granularity. To avoid political risks in the early stages of an open election data project, an EMB may also want to start with the publication of politically less sensitive data.

For prioritizing data sets, much depends on the context, but results data, voter registration data, legislative, district boundaries and political finance data may well be high up on the priority list. Here, stakeholder involvement is key to gaining an understanding of their interests and of the scale of the demand for data.

Selecting an open-data licence

Legal openness, i.e. making sure that published data is not subject to any copyright or other usage restrictions, is a key principle of open election data. This openness must be clarified and communicated to potential users through an explicit open licence under which the data is made available.

Some countries and governments will already have open government licences in place that can also be used for electoral purposes. The UK's Open Government License (OGL) is one example of such a standardized licence (National Archives n.d.).

Where such a government-wide open data licence is not available, Open Data Commons (n.d.) and Open Knowledge International n.d. (d)) recommend several options, including:

- a Creative Commons License, which is widely used and compatible with most other licences;
- an Open Data Commons Public Domain Dedication License (PDDL), which waives all rights;
- an Open Data Commons Attribution License (ODC-BY), which requires attribution; or
- an Open Data Commons Database License (ODbL), which requires users to share and attribute similar to the Creative Commons Attribution ShareAlike (CC BY-SA) licence.

Applying a licence only requires the addition of a statement on the page where the data is made available. Examples of such statements include:

- for PDDL:

  This data is made available under the Public Domain Dedication and License v1.0, its full text can be found at: <http://www.opendatacommons.org/licenses/pddl/1.0/>.

- for the ODC-BY:

  This data is made available under the Open Data Commons Attribution License: <http://opendatacommons.org/licenses/by/1.0>.

- for the ODbL:

  This data is made available under an Open Database License, whose full text can be found at <http://opendatacommons.org/licenses/odbl/>. Any rights to the individual contents of the database are licensed under the Database Contents License, whose text can be found at <http://opendatacommons.org/licenses/dbcl/>.

Publishing data

The publishing step is where many of the open-data principles need to be technically implemented. The technical infrastructure ensures that the data is publicly available, timely, permanent, free of charge, without any restrictions, in bulk and in a machine-readable and open format.

Timely and permanent data release requires data to be available to the public as soon as possible and certainly before it becomes irrelevant. This will be different for each data set. Whether or not quality control and data audits can and should be conducted prior to the first publication is an important decision: on the one hand, mistakes should be corrected before publication in order to not raise stakeholder doubts regarding the quality of the data. If such quality checks take too long, however, they will defeat the goal of timely data release.

The permanent availability of data entails assurances that the data will be available on a stable, long-term data portal that will neither crash due to the high degree of usage on election day nor be taken offline after the electoral period.

File and data formats

Providing data in an open, machine-readable format ensures that data users can process the data efficiently and are not required to purchase any proprietary technology to access the data. Open and machine-readable formats include, for example, comma-separated values (CSV), tab-separated values (TSV) or JavaScript Object Notation (JSON).

PDF documents are not considered machine-readable. This format is suitable for displaying and printing information, but it makes automated data analysis either very difficult or completely impossible.

In countries with a decentralized election administration, it is important to consider standardizing the format in which data is published by the various competent bodies. If each administrative body uses even a slightly different format, combining and using data across the country will become more difficult and error-prone. In the United States, for example, a common data format for reporting election results was established by the National Institute of Standards and Technology Voting System Standards Committee (Wack 2014). Such standards ensure that data provided by electoral officials in different jurisdictions is easier for the general public to export, import and utilize.

The Open Data Institute (Dodds 2014) has developed a data standard for election data with an international perspective.

Online publication

While data has been published using many different media, open election data is commonly expected to be published online for the most widespread access possible. Where Internet access is limited for some stakeholders or where the data sets in question are extremely large, additional publication on storage media may be a consideration.

For online publication, several different options are available:

- **Directly on the EMB's website**: This is where interested stakeholders will likely to go first when looking for election data, and as no data transfer between institutions is necessary, this may technically be the easiest solution. The disadvantage of publishing on the EMB's website can be that it may not have all the features and formats required for open data publication and that the website may possibly even crash

on election day. In case a new functionality for the publication of data needs to be created, there is no need to build the required software from scratch. Free and open-source software solutions such as those produced by the Comprehensive Knowledge Archive Network (CKAN) are available and provide ready-made data management platforms for making data accessible through tools that allow integrating, publishing, sharing, finding and using data.

- **Government open-data portals:** some countries have local, regional or national open-data portals that provide access to all kinds of government data. Such portals may also be used for election data. The advantages of this approach include that users can intuitively find election data where they may already access other data sets. Established open-data portals may also be more reliable and better maintained, well tested and have a publication functionality that does not need to be created from scratch. Third-party sites can also offer a community of interested people and other data. Utilizing existing open-data portals has the advantage that the operators of such portals will have experience that the EMB can tap into. A downside is that efficient data exchange between the EMB and the open-data portals needs to be established, especially on election day. This includes establishing mechanisms for exchanging data in real time.

When publishing data, the objective is to maximize access and use. Instead of choosing between posting data on a national portal and on an institution's own website, one option is to host the data on a single platform and then direct users of other websites and portals to it. These secondary portals or websites can act as catalogues, providing access to the data through a link. Such an approach provides multiple opportunities for the data to be found, while any updates to the data are reflected across all the sites where it is presented. The UK Government's national portal is, for example, in effect a catalogue that harvests election data from other sites, including the Electoral Commission's website (Data.gov.uk 2016; Electoral Commission 2016(a)).

Bulk download of raw data vs. APIs

No matter which location is selected for data publication, there are two different methods for providing the data. Availability in bulk is important, as it allows users to download data as one complete set and not as a cumbersome sequence of many small downloads.

Data can be provided as data tables or files for bulk downloading of raw data. Bulk downloads are usually transferred using HyperText Markup Language (HTML), the standard format for website data, or, less commonly, using File Transfer Protocol (FTP). Both formats make it possible to download full data files in formats such as CSV.

The second option for data access is via an API, a format that allows the automatic selection of specific parts of data instead of downloading entire data sets. This format can be very useful for real-time publication of very large data sets. APIs can, for example, eliminate the need for data consumers to continuously download entire data sets to get the latest data; instead, they can detect and access only those parts that have been updated since their last download. APIs are common for data that is updated frequently, such as transportation and weather data.

While APIs can be efficient, they are recommended only as an additional access method in addition to downloading files for several reasons. For example, APIs are more costly to develop and maintain, they are also more difficult for data consumers to use and, most importantly, they are transient and cannot be archived as easily as data files.

Testing and using open data sources

Testing open-data portals from a user's point of view reveals that many systems have surprising weaknesses that make it seem as though the publishing institution has rarely tried to utilize its data interface themselves. It is therefore recommended to also try using the provided data or to actively engage with the data user community to learn about such difficulties.

Make data discoverable

Election data is of great interest to many electoral stakeholders, and it can be expected that they will actively search for, and demand access to, this data. It is important that open election data be easily discoverable.

Promoting the availability of open election data will allow an EMB to communicate the steps it has taken to increase trust in, and the transparency of, the institution and its processes. Additionally, widely publishing sources of election data ensures that all stakeholders have equal opportunities to utilize the data and that no stakeholders are privileged by knowing how to locate otherwise concealed information. Access to open election data should also be equitable because privileged access for media, officials or researchers can undermine community participation and engagement.

Joining forces with other governmental publishers can be beneficial and may make it easier to discover data, but it is important that inter-agency politics, future budget cycles, jurisdictional borders, whether sectoral or geographical, should not make cooperation difficult (Open Knowledge International n.d. (g)).

As mentioned above, many governments have created government data catalogues that may already have space for the inclusion of election data, and collaborating with such catalogues increases the visibility of the published data.

CHAPTER 6

Conclusions and recommendations

1. Open data initiatives aim at improving electoral integrity by increasing transparency, addressing information asymmetries and granting more equal power to electoral stakeholders.

2. Open data is an opportunity for EMBs to demonstrate a commitment to transparency and accountability.

3. Some case studies exist concerning the impact of open data on electoral processes; however, relatively little election data is currently available openly, and all electoral stakeholders can do more to rectify this.

4. Open-data strategies can be based on global principles and recommendations, but they also need to be contextualized to the specific priorities of the stakeholder and the country.

5. There are a number of risks that should be understood and mitigated against when designing any open-data strategy, including around privacy and the misuse of data.

6. Even if data publication according to all open-data principles is not possible, publishing data in the best available format can still unleash the potential inherent in open data.

7. Open data requires the establishment of legal and technical building blocks, but it is most impactful when it entails a cultural change at the EMB.

8. An EMB does not necessarily need to start from scratch with open data: initiatives and expertise may already exist and may be accessed through other parts of government, in local civil society and in global initiatives and partnerships.

9. Close cooperation with relevant electoral stakeholders increases the effectiveness of open-data initiatives, as it increases the likelihood that useful data will be published and that these stakeholders will be able to use it.

10. Data is not necessarily user-friendly for non-experts. Tools that make data easier to engage with, as well as intermediaries such as journalists or civil technologists, may be needed to process the data to make insights accessible to a wider audience.

11. Open data is only one element that can add to trust and transparency in elections; however, open data cannot compensate for or replace shortcomings in other areas of the electoral process.

References and further reading

5 Star Open Data, 2012, <http://5stardata.info/en/>, accessed 15 December 2016

ACE Project, '2014 Indonesian Elections Case Study', <http://aceproject.org/ace-en/topics/et/ety/ety07>, accessed 11 November 2016

Afghanistan Election Data, <http://afghanistanelectiondata.org/>, accessed 11 November 2016

Bardall, G., 'Engendering Election Data', Presentation at the Electoral Risk Management Conference Addis Ababa, 2015, <http://www.idea.int>, accessed 20 October 2016

Berners-Lee, T., 'The next web', TED Talk, 2009, <https://www.ted.com/talks/tim_berners_lee_on_the_next_web>, accessed 15 December 2016

Berners-Lee, T., Hendler, J. and Lassila, O., 'The semantic web', *Scientific American*, 284/5 (2001), pp. 28–37

Carolan, L., *Open data, transparency and accountability: Topic guide* (Birmingham, UK: Governance and Social Development Resource Centre, University of Birmingham, 2016(a))

—, 'Open elections: lessons for open data from Burkina Faso and Zambia', Open Data Institute, 10 May 2016(b), <https://theodi.org/blog/open-elections-lessons-for-open-data-from-burkina-faso-and-zambia>, accessed 14 November 2016

Data.gov.uk, 'EU referendum results, 2016, <https://data.gov.uk/dataset/eu-referendum-results>, accessed 16 December 2016

Development Seed, 'Afghanistan Elections: Open Data Helps Afghans to Demand Better Elections', <https://developmentseed.org/projects/afghanistan-elections/>, accessed 15 December 2016

—, 'Afghanistan Elections: Why We Opened Afghanistan's Election Data: Better Data for Better Elections', 3 March 2014, <https://developmentseed.org/blog/2014/03/03/why-afghanistan-open/>, accessed 15 December 2016

Dodds, L., 'Research: investigation into publishing open election data', Open Data Institute, 8 August 2014, <https://theodi.org/blog/election-data-tables>, accessed 15 December 2016

Doyle, J., 'Electoral roll data leaked: Millions of names and addresses may have been illegally sold to junk mail firm', MailOnline, 9 May 2014, <http://www.dailymail.co.uk/news/article-2623853/Electoral-roll-data-leaked-Millions-names-addresses-illegally-sold-junk-mail-firm.html>, 16 December 2016

Electoral Commission, Electoral Data, 2016(a), <http://www.electoralcommission.org.uk/our-work/our-research/electoral-data>, accessed 15 December 2016

—, 'Investigation in respect of the Labour Party 2015 UK Parliamentary General Election campaign spending return', 25 October 2016(b), <http://www.electoralcommission.org.uk/__data/assets/pdf_file/0003/216507/Labour-UKPGE-Investigation-Report.pdf>, accessed 15 December 2016

Foti, J., 'Aligning Supply and Demand for Better Governance: Open Data in the Open Government Partnership', Open Government Partnership Blog, 27 May 2015, <http://www.opengovpartnership.org/blog/joe-foti/2015/05/27/aligning-supply-and-demand-better-governance-open-data-open-government>, accessed 14 November 2016

Gowda, A., 'Karnataka government risks women safety, puts out personal data in caste census', *India Today*, 9 July 2015, <http://indiatoday.intoday.in/story/karnataka-government-caste-censuswomen-privacy-violated/1/450140.html>, accessed 14 December 2016

Graft, A., Verhulst, S. and Young, A., 'Indonesia's Kawal Pemilu—Elections: Free, Fair and Open Data', January 2016, <http://odimpact.org/case-indonesias-kawal-pemilu.html>, accessed 14 November 2016

International Foundation for Electoral Systems, 'Ethiopia: Implications of the May 2005 Elections for Future Democratization Programs', August 2005, <https://www.ifes.org/sites/default/files/ifes_ethiopia_dg_options_paper.pdf>, accessed 15 December 2016

International IDEA, *The Use of Open Source Technology in Elections* (Stockholm: International IDEA, 2014(a))

—, *Electoral Management Design: Revised edition* (Stockholm: International IDEA, 2014(b))

—, ICTs in Elections Database, <http://www.idea.int/data-tools/data/icts-elections>, accessed 14 November 2016

Kaplan, J., 'Is Data Undermining Our Democracy?', *Huffington Post*, 5 August 2016, <http://www.huffingtonpost.com/entry/is-data-undermining-our-democracy_us_57a47d89e4b034b25894b06b>, accessed 16 December 2016

Larrick, S., 'Shouldn't text be open data too? The search for an inclusive data definition', 10 August 2016, <https://sunlightfoundation.com/2016/08/10/shouldnt-text-be-open-data-too-the-search-for-an-inclusive-data-definition/>, accessed 14 November 2016

Mourakiboun, 2014 Parliamentary Elections, <http://tunisiaelectiondata.com/>, accessed 14 November 2016

National Archives, Open Government License for public sector information, <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/1/>, accessed 14 November 2016

National Democratic Institute and Google, Open Election Data Initiative, 2015(a), <http://www.openelectiondata.net/>, accessed 1 October 2016

—, Latin America Data Inventory, 2015(b), <http://www.openelectiondata.net/en/inventory/>, accessed 14 November 2016

—, Open Election Data Principles, 2015(c), <http://www.openelectiondata.net/en/guide/principles/>, accessed 14 November 2016

Open Data Charter, 2015, <http://opendatacharter.net/>, accessed 17 November 2016

Open Data Commons, 'Making Your Data Open: A Guide', <http://opendatacommons.org/guide/>, accessed 14 November 2016

Opendata.go.tz, 'Tanzania Government Open Data Portal', <http://www.opendata.go.tz/>, accessed 15 December 2016

Open Data Institute, 'Open data in government: how to bring about change', 2015(a) <http://theodi.org/open-data-in-government-how-to-bring-about-change>, accessed 15 December 2016

—, Open Data Maturity Model, 2015(b) <http://theodi.org/guides/maturity-model>, accessed 15 December 2016

—, 'How to write a good open data policy', 2016, <http://theodi.org/guides/writing-a-good-open-data-policy>, accessed 16 December 2016

—, Open Data Pathway n.d. (a), <http://pathway.theodi.org/about>, accessed 15 December 2016

—, The Data Spectrum n.d. (b), <http://theodi.org/data-spectrum>, accessed 14 November 2016

Open Data Research Network, 'Open Data & Privacy - discussion notes', 2013, <http://www.opendataresearch.org/content/2013/501/open-data-privacy-discussion-notes>, accessed 14 November 2016

Open Government Partnership, Open Government Declaration, 2011, <http://www.opengovpartnership.org/about/open-government-declaration>, accessed 14 November 2016

—, Participating Countries, 2016, <http://www.opengovpartnership.org/countries>, accessed 15 December 2016

—, 'Open Government Guide: Privacy and data protection', <http://www.opengovguide.com/topics/privacy-and-data-protection/>, accessed 15 December 2016

Open Government Working Group, 'Request for Comments: Open Government Data Principles', 7-8 December 2007, <https://public.resource.org/8_principles.html>, accessed 14 November 2016

Open Knowledge International, Data Portals: A Comprehensive List of Open Data Portals from Around the World n.d. (a), <http://dataportals.org/>, accessed 14 November 2016

—, 'Open Data Handbook' n.d. (b), <http://opendatahandbook.org/>, accessed 14 November 2016

—, Open Definition n.d. (c), <http://opendefinition.org/about/>, accessed 14 November 2016

—, Open Definition: Conformant Licenses n.d. (d), <http://opendefinition.org/licenses/>, accessed 14 November 2016

—, Global Open Data Index n.d. (e), <http://index.okfn.org/>, accessed 14 November 2016

—, Global Open Data Index: Election Results n.d. (f), <http://index.okfn.org/dataset/elections/>, accessed 14 November 2016

—, How to Open up Data n.d. (g), <http://opendatahandbook.org/guide/en/how-to-open-up-data/>, accessed 14 November 2016

—, The Open Knowledge Network n.d. (h), <https://okfn.org/network/>, accessed 15 December 2016

—, What is Open? n.d. (i), <https://okfn.org/opendata/>, accessed 14 November 2016

—, Why Open Data? n.d. (j), <http://opendatahandbook.org/guide/en/why-open-data/>, accessed 14 November 2016

Robinson, D., Yu, H., Zeller, W. and Felton, E., 'Government data and the invisible hand', *Yale Journal of Law & Technology*, 11/1 (2009), pp. 60–75, <http://digitalcommons.law.yale.edu/yjolt/vol11/iss1/4/>, accessed 15 December 2016

Schuller, I., 'Tunisia Open Election Data: Mourakiboun makes Tunisia election data open and accessible', 22 May 2014, <https://developmentseed.org/blog/2014/05/22/tunisia-open-election-data/>, accessed 14 November 2016

Scott, A., 'Case study: Burkina Faso's open elections', <http://theodi.org/case-study-burkina-fasos-open-elections>, accessed 14 November 2016

Sunlight Foundation, 'Ten Principles for Opening Up Government Information', 11 August 2010, <https://sunlightfoundation.com/policy/documents/ten-open-data-principles/>, accessed 14 November 2016

Transparency and Accountability Initiative and Open Government Partnership, *Open Government Guide,* <http://www.opengovguide.com/>, accessed 14 November 2016

Ubaldi, B. 'Open Government Data: Towards Empirical Analysis of Open

Government Data Initiatives', *OECD Working Papers on Public Governance*, No. 22, 2013 <http://www.oecd-ilibrary.org/governance/open-government-data_5k46bj4f03s7-en>, accessed 14 November 2016

Wack, J., 'IEEE VSSC/1622: Voting System Standards', *IEEE Computer*, 47/9 (2014), pp. 94–97, <https://www.nist.gov/node/560811>, accessed 17 November 2016

Whong, C., 'Thoughts on Open Data', 2015 <http://chriswhong.com/open-data/thoughts-on-open-data-part-1>, accessed 14 November 2016

World Bank, Data Readiness Assessment Tool, <http://opendatatoolkit.worldbank.org/en/odra.html>, accessed 15 December 2016

World Wide Web Consortium (W3C), Linked Data, <https://www.w3.org/standards/semanticweb/data>, accessed 15 December 2016

Yu, H. and Robinson, D. 'The new ambiguity of "open government"', *UCLA Law Review Discourse*, 59 (2012), pp. 178–208, <http://www.uclalawreview.org/the-new-ambiguity-of-%E2%80%9Copen-government%E2%80%9D>, accessed 15 December 2016

About the contributors

Liz Carolan supports governments and teams in using data and transparency to achieve their objectives, both for the Open Data Institute (ODI), where she is an associate, and as an independent advisor. In this capacity, she has advised leaders from over 30 countries, including through setting up the Open Data Leaders Network, and she led the ODI's work that supported the first-ever publication of real-time election results in Burkina Faso. Among her previous roles, she developed learning programmes for politicians and civil servants at the Institute for Government and the Africa Governance Initiative. Her prior publications include 'Open data, transparency and accountability: Topic guide' published by the GSDRC and the United Kingdom's Department for International Development.

Peter Wolf is a member of the Elections Processes Programme at International IDEA. He has focused on ICT applications in electoral processes since joining the Elections Department of the OSCE mission in post-war Bosnia and Herzegovina, where he worked on voter registration and results databases. He has served as a consultant in voter registration projects in Albania, the Democratic Republic of the Congo and Iraq, and he has participated in various international election observation missions, including as an electronic voting expert in France, Kazakhstan, Kyrgyzstan, the Philippines and Venezuela. He holds a master of science degree from the Graz University of Technology.

About International IDEA

The International Institute for Democracy and Electoral Assistance (International IDEA) is an intergovernmental organization that supports sustainable democracy worldwide. International IDEA's mission is to support sustainable democratic change by providing comparative knowledge, assisting in democratic reform, and influencing policies and politics.

What does International IDEA do?

In the fields of elections, constitution-building, political parties, gender in democracy and women's political empowerment, democracy self-assessments, and democracy and development, International IDEA works in three main activity areas:

- providing comparative knowledge derived from practical experience on democracy-building processes from diverse contexts around the world;
- assisting political actors in reforming democratic institutions and processes, and engaging in political processes when invited to do so; and
- influencing democracy-building policies through the provision of comparative knowledge resources and assistance to political actors.

Where does International IDEA work?

International IDEA works worldwide. Based in Stockholm, Sweden, it has offices in Africa, Asia, Europe and Latin America.

The concept of open data is based on the idea that data should be freely available for anyone to access, use and share. In recent years, the concept has taken root as an increasingly accepted practice for public and government produced data, with data seen as a public good, and data infrastructure as the key to accessing it.

Publishing election data as open data has numerous benefits: it provides civil society, citizen journalists, electoral observers and citizens access to the same detailed information that was previously only available to selected stakeholders such as large media outlets. In doing so, open data allows all interested stakeholders to follow and understand the electoral process and can enable more inclusive, transparent and trusted elections.

In spite of the potential benefits of open data, election data is often not available in government open-data platforms and election data that is published often does not comply with open-data principles.

The aim of this publication is to encourage electoral management bodies to increase the application of open-data principles. It provides guidance for improving the practice of data publication towards a more open approach.