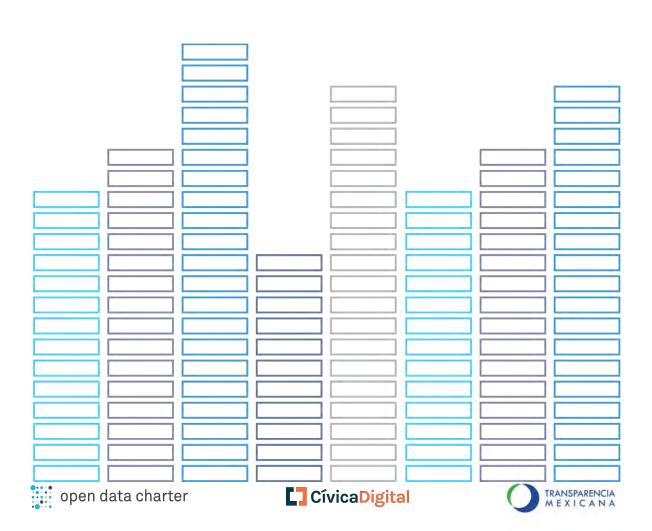


# **Open Up Guide**

# Testing how to use open data to combat corruption in Mexico

January 2018

#### **CONTENT**



Summary	3
Our Approach	4
What did we find?	5
Mapping anti corruption data: Data readiness	6
Horizontal evaluation: Data consistency	6
Vertical evaluation: Data quality	7
Recommendations and next steps	8
Annex 1	12

## Acknowledgements

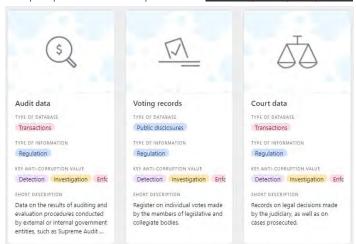
This report was written by Alejandro Echeverria (Open Data Charter), Diana D'Herrera and Ricardo Alanís (Cívica Digital). It was edited by Ania Calderon and Robert Palmer (Open Data Charter). We would like to thank Enrique Zapata, Andrea Barenque, Nancy Camacho, Dante Preisser and Alejandra Rascón from the Government of Mexico, and Eduardo Bohorquez, Lucía Petersen and María Zilli from Transparencia Mexicana for their support in delivering the project. The project was funded by the Inter-American Development Bank, the William and Flora Hewlett Foundation and the Omidyar Network.

### Summary

The Open Data Charter has collaborated with the government of Mexico, Transparencia-Mexicana and Cívica Digital, to test how open data can be used to combat corruption in Mexico. The project involves implementing the Charter's <a href="Anti-Corruption Open Up Guide">Anti-Corruption Open Up Guide</a>.

When money that should be spent on government services ends up in the hands of dishonest officials, it can have a damaging impact on political, economic and social spheres. Corruption is driven by networks of individuals, professional intermediaries and organised structures. In order to tackle this crime, it's important to understand, identify and dismantle such networks, and to have legislation and preventive systems that discourage their formation. Corruption schemes frequently rely upon the law to secure ownership of companies, land and assets used to launder their proceeds. Opening up and using, government datasets can provide the information needed to prevent and fight corruption.

At the heart of the Anti-Corruption Open Up Guide is <u>a list of the top 30 types of dataset</u> that can be used for anti-corruption purposes, along with the features that each dataset needs to have in order to be effective.



Example of Dataset Cards as found in the Anti-corruption Open Up Guide.

This implementation project had three objectives:

- Audit which datasets are already collected and published by the government of Mexico.
- Provide recommendations on how to improve publication and use of anti-corruption related data.
- Update the Guide based on findings on the ground.

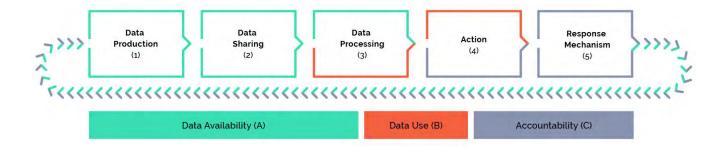
Through assessing the datasets that the government already releases, and interviewing government officials that generate, use or publish data, the project has developed a better understanding of the state of relevant open data in Mexico. We have produced a set of recommendations for how to improve data publication:

• Develop official governance frameworks for key anti-corruption data such as

- o a national inventory of anti-corruption data systems
- o documentation of production, quality assurance and updating processes,
- o appointment of custodians for data systems and registers (areas responsible of the management of each dataset), and
- o development of open APIs to access and share the datasets.
- Follow internationally agreed open standards across datasets as recommended by the <u>Open Up Guide.</u>
- Improve the interoperability between key information systems and datasets, as well as to promote human and institutional interoperability by getting officials within different government offices to speak to each other, reduce duplication and share best practice.
- Encourage communication between data producers in government and data users in civil society, the media, business and journalism, focused on the use of data for addressing anti corruption policy challenges.
- Recommend best practices to be included as part of the work of the National Anti Corruption System and its digital platform, including engaging the legislative and judiciary branches for datasets produced under their mandates.

## Our Approach

The starting point was based on our understanding of how transparency can lead to accountability. In a <u>paper published</u> in June 2017 in collaboration with the Transparency and Accountability Initiative, we set out our initial ideas on how to get impact from open data. Although we recognise that this process is never as linear as the diagram below suggests, we think it's helpful to think about the flow from data generation, to use, to action and reaction.



The project in Mexico allows us to test some of our assumptions in practice. During this phase we tested the following assumptions:

- 1. That the datasets in the Guide can be collected and **produced** by government.
- 2. That the datasets can be **shared** by government in the ways the Guide advises.
- 3. That a subset of datasets can be prioritized for use and **action** to tackle key anti corruption challenges.

A key challenge for the project was to define what good open data looks like for each of the datasets. The Guide includes recommendations of which standards to use for many of the types of datasets. However, the production of data often does not follow general standards because

the responsibles of generating and managing the data are frequently tasked to manage specific table or dataset limited to their functions.

In order to get a better picture of the anti corruption data landscape in Mexico, we evaluated the available datasets against three main criteria:

- Readiness: Does data follow basic Charter principles, (ie, Licensing, data update interval, bulk data availability)? This work drew on the methodology of the <u>Open Data Barometer</u> and the <u>Global Open Data Index</u>. (See <u>this blog</u> for more information on the different open data measurement tools).
- 2. **Consistency:** Was the data found without empty rows or errors on data structure? This was a "horizontal evaluation", as it is relates to the horizontal axis of data, row by row.
- 3. **Quality:** To what extent, does the dataset comply with the data standards? This was an evaluation of the data 'dimensions' and 'features' as understood in data science. It looked at the extent to which the dataset included the necessary fields for the type of data.

#### What did we find?

Initial findings revealed a mixed picture. The government of Mexico has already made efforts to publish data that can be used to fight corruption, but the challenge is to increase its quality to meet international standards. However, by speaking to officials involved in data production, sharing and use, we were able to get some important insights.<sup>1</sup>

- 1. By and large the government officials we spoke to were committed to improving the collection, publication and use of data. They talked frankly about the problems they faced and were interested in collaborating with us and our civil society partners in getting better data out there and improving their own data literacy skills.
- 2. Datasets often lack governance frameworks and are not interoperable. One of the recurrent themes throughout the sessions with officials was that existing systems and datasets in different government offices were not intended to be open or built to communicate with each other. Hence, there is a need to identify points of contact across systems to be able to build internal processes that are efficient and transparent across board.
- 3. There were some commonly used identifiers across different datasets (e.g. for individuals or companies) but due to the siloed nature of the production of data there was no coordination around this. A shared set of identifiers is crucial to make datasets produced by different bits of government inter-operable. For example, sharing identifiers of sanctioned companies with the register of government contractors could prevent barred companies from obtaining government bids.
- 4. **It's not just datasets that need to be to be able to talk to each other people do as well.** We carried out our interviews in small groups of officials doing similar work. It was striking how many of them had never met before, and didn't know that someone else in government was doing similar work.
- 5. There were examples of data being produced and shared, but often it was hard to find where. Files are given long technical names or use acronyms that only policy experts understand. Accessibility would improve if the government simplified names and used language that ordinary citizens can understand.

<sup>&</sup>lt;sup>1</sup> These are drawn on two blogs published immediately after our interviews with government officials in August 2017. See <a href="here">here</a> and <a href="here">here</a>.

- 6. Most government offices lacked the technical capacity, systems and other resources needed to process large amounts of data. Some public offices also depended on renewing software or service licenses to carry out their work, which in turn hampers their internal capacity.
- 7. While some data publishers were talking to data users, there was no systematic way of doing this across government. Part of our aim with this project is to get this to happen early on in the process so that producing and sharing data is done with an end user in mind. Greater coordination with access to information requests and maintaining regular engagement with data user groups should help to improve the quality of the data and increase its usefulness.
- 8. A big barrier to publication of data is concerns around privacy, data protection and secrecy laws. While the Charter's principles recognise the fundamental importance of respecting privacy, this needs to be balanced with the commitment to be "open by default". In Mexico, as in other jurisdictions, secrecy concerns can sometimes be used as a smokescreen to not publishing public information at all.

#### Mapping anti corruption data: Data readiness

Building on these insights, we carried out a data audit using a series of sprints to analyse the data sources, with the intention of finding the most important challenges related to open data *readiness, consistency and quality,* and the potential of the datasets to be used for the combat of corruption.

The full initial mapping, carried between September and November of 2017, can be found in the <a href="Extended Mapping"><u>Extended Mapping</u></a> Annex (in Spanish).

Our key takeaways on basic data readiness were:

- The 30 key general datasets recommended in the Guide, translated into 72 specific datasets in the Mexican context, 3 were in the process of being released, with 47 already available in the national open data repository, datos.gob.mx and 22 which do not cover basic data readiness and/or are not published. (See Annex 1)
- 100% of the data sources found on datos.gob.mx were machine readable, while only 45% of the data sources not found on datos.gob.mx were machine readable.
- 33% of the data sources found could be downloaded in bulk.
- Data on datos.gob.mx has a de-facto open data license (libre uso mx) and is DCAT dictionary data. There is a citizen backup (datamx.io), with an API using CKAN, helping improve its access
- Approximately 75% of the data sources were found to be updated in a reasonable window.
- By being in an open data system, or on datos.gob.mx, 80% of the data sources were relatively easily to find.
- Out of the 30 datasets listed in the Guide we found that the government was publishing some relevant data on 19 of them, out of which four where not available as open data, consisting Government Council or Advisory Boards Register, Meeting Records, Land Register and Public Procurement Complaints Register.
- Of the newly identified datasets that were not open, five are in the process of being
  published by the Secretaria de Función Pública (SFP) and the Secretaria de Hacienda y
  Crédito Público (SHCP), and correspond to PPPs, Audit Data and Government Grants'
  datasets.

• There is a lack of culture of implementing concrete data governance frameworks and documenting process of data production.

#### Horizontal evaluation: Data consistency

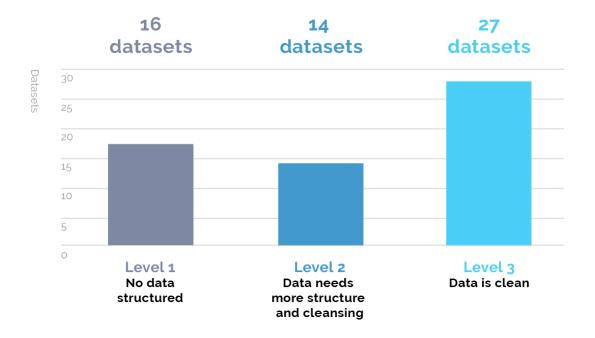
In the horizontal evaluation, looking at data consistency, data sources were rated against three levels:

- Level 1 -where no data is structured:
- Level 2 -where data needs more structure and cleansing, and;
- Level 3 where data is relatively clean. This was done by testing entity consistency through the datasets, correct use of data separators and overall technical quality of the dataset.

Our key takeaways on data consistency were:

- The most consistent reason to rate a low quality data register was that data was not found to be structured. For the intermediate level, the main problem was that particular government entities were represented in a number of different ways, making it hard to track them across the register.
- Data found at datos.gob.mx was generally found to be more consistent than data sources found outside the central Mexican repository.
- As the Guide can be implemented by different data generators, Legislative and Judiciary
  Powers would benefit from implementing open data principles in similar fashion to offices
  that took part in this exercise.

Graph 1: Distribution of datasets in relation to data consistency (the horizontal evaluation of the initial set of analyzed data).



#### Vertical evaluation: Data quality

In the vertical evaluation, looking at how well the datasets met the relevant standards, the data sources were also rated against three levels, according to the qualitative difference between the dataset as published, taking into account the dimensions of both sides. Data was found to be skewed to the middle to top level, with only some datasets to be found in the lower end of the spectrum.

Our key takeaways on data quality:

- The most consistent reason to rate a data source as of low quality was that it did not abide by the defined standards.
- Challenges to cover as much of the international data standards were encountered both by data found outside and inside the national Mexican repository, regardless of whether there was an intention to cover the standard or not.

Graph 2: Distribution of datasets in relation to data quality (the vertical evaluation of the initial set of analyzed data).



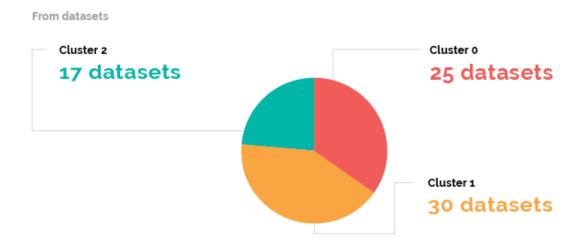
# **Recommendations and next steps**

From the evaluation, some patterns emerged. By assigning grades to points assessed by the evaluation, three major clusters of datasets were identified:

• Cluster 0: Data sets with fundamental challenges that come from the readiness assessment, such as the frequency of data upload and licensing issues. Most of the issues are solvable by becoming part of datos.gob.mx. Datasets in this category were related to Campaign Promises and Changes in Regulations.

- Cluster 1: Datasets with issues on data consistency which are solvable with a solid data pipeline that removes mistakes and solidifies data availability. The most recurring datasets in this level were Political Party Finance, Audit Data and Changes in Regulation. Additionally, it was noted that the public offices that hold the largest amount data in this level are the Secretaria de Función Pública (SFP) and the Secretaría de Hacienda y Crédito Público (SHCP) given the national mandate of both institutions of carrying out activities such as coordinating public spending, tax collection and public exercise.
- Cluster 2: Good datasets with some opportunities to improve their quality. That is, data that is clean and follows the Open Data Charter principles, so it can be pushed towards covering more requirements of the relevant standards. At this level we noted a more spread out sample of datasets ranging from Company Register, Contracts Register, Court Data, Audit Data and Charity Register. Most datasets were generated within offices at the Secretaría de Función Pública (SFP) and the Secretaría de Hacienda y Crédito Público (SHCP).

Graph 3: Level of readiness and quality of anti corruption data in clusters



\*For a complete list of data sources identified in each cluster see Annex 1

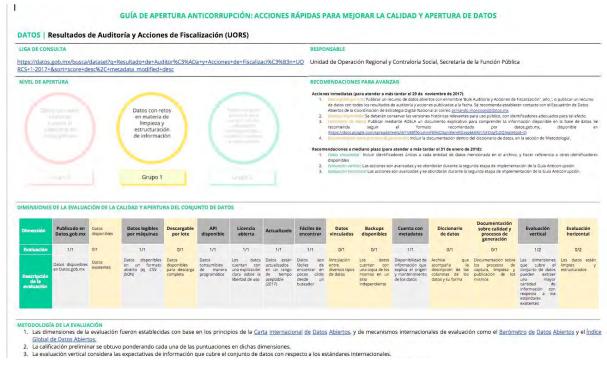
Categorizing data sources according to their level of openness, allowed us to design a process to prioritise actions to improve their accessibility. This approach will allow government officials who produce and share data to improve the quality of the data they're providing, following international standards that can enable their usability. Further targeted actions were recommended for the use of datasets that ranked the highest in levels of readiness, consistency and quality:

- **Cluster 0**: Open data principles bootcamp training. This group should work towards the basics, such as understanding the value of an open data license and sharing their data in bulk
- **Cluster 1:** Open data pipeline improvement and data cleansing. This group should work towards a more structured data processing and sharing strategy.

• **Cluster 2**: Anti-corruption Open Up Guide training. As the basic challenges of open data have been accomplished, this group is ready to enrich the datasets based on user feedback and priority policy challenges.

In order to follow up the evaluation of each dataset and to generate a clear dialogue with government and with civil society, the government of Mexico generated a public user-friendly record that contains the following for each dataset:

- Name of the register, data owner and source of the data evaluated. (Heading)
- Cluster of the register and the challenges ahead. (Second row, left cell)
- Detailed evaluation of the register, described point by point. (Lower row)
- Top recommendations to the data owner using a non technical vocabulary in line of the data source challenges on data readiness, consistency and quality. (See each assessment card below for specific recommendations). E.g. "Available in Bulk Adding a dataset that includes all the audits."



Example assessment card noting cluster and recommendations to communicate the challenges ahead.

Based on these recommendations, datasets were published in the National Open Data Portal, and grouped under a simplified search for <u>Anti-corruption Data Infrastructure</u>. The next stage of the pilot project will involve setting up a governance structure to support the management of an anti corruption data infrastructure as part of the National Anticorruption System, and working with government and civil society to use the datasets found in Cluster 2 to help tackle a specific aspect of corruption. This will test the ability of open data to be processed by anti corruption advocates and to promote and institutionalise action by government.

Graph 4: Cluster 2 datasets ranking highest in readiness and quality

Cluster	Public Institution	Related Dataset from the Guide	System/Register/Database
2	COFEMER -	Records of changes in regulations	Regulatory Draft sent by the agencies to COFEMER
	INDESOL -	Charity Register—	— Federal Register of Civil Society Organizations (OSC)
	INEGI	Company Register —	National Statistical Directory of Economic Units (DENUE)
	PGR —	Court data	Criminal incidence of federal jurisdiction
			<ul> <li>Statistics of the General Direction of Control of Criminal Proceedings and Appeals in the Matter of Federal Crimes (DGCPPAMDF)</li> </ul>
	SAT —	Charity register ————————————————————————————————————	Directory of Authorized Grantees
	SFP —	Debarred or sanctioned contractors	Directory of sanctioned Suppliers and Contractors (DPCS)
		– Audit Data –	Results of Appointment of External Auditors
	SHCP	PPPs	Registration of Pure Public-Private Partnerships
			Registration of Projects of Public-Private Partnerships
			Registration of the investment work portfolio
		Spending —	Assignment and execution of the budget of expenses of the federation
			— Federal entities projects
			Federalized spending contracts
	SSA —	- Charity register	Amounts granted in joint venture projects with civil society organizations
	SE —	International aid and financing	List of the Registry of Foreign Companies registered in the National Registry of Foreign Investments that perform acts of commerce in Mexico
	SEP -	Lobbying register	- Open Contracting

The Anti-Corruption Open Up Guide will be updated as more datasets are identified in addition to the ones made available. Changes will be reflected through the <u>online resource</u> and illustrated as seen with the existing 30 dataset cards.

#### Annex 1

Mapping of anti corruption data in Mexico

# Table 1: Clustering of data sources according to their level of readiness, consistency and quality

Cluster	Public Institution	Related Dataset from the Guide	System/Register/Database
0	CD —	Voting records	Legislature voting
	CS —		Voting carried out during the legislatures
	INE	Political parties finances	— Integral Audit System
			Lobbying Register
			Historic Register of Public Financing to Political Parties
			Yearly Reporting on Political Parties Spending
		- Campaign promises ———————————————————————————————————	— Political Parties Prorgammatic Platform
	PGR —	Court data	List of Previous Investigations
			— International Bribery
	SE	Company Register —	List of Foreign Companies within the National Investment Register
	SEDATU	- Land register	National Platform for Register and Cadastral Information
	SEGOB	Records of changes in regulations ————	— Official Journal of the Federation (DOF)
		Council / advisory board members ———	— Participatory Mechanisms Data Base
	SFP —	- Debarred or sanctioned contractors ————	— System of Administrative Procedures of Sanction to Suppliers and Contractors
	SFP	- Audit data	Evaluation of the Performance of Signatures of External Auditors
			Number of Partners removed
			Institutional Internal Control System (SCII)
		Licenses, concessions and permits ———	List of procedures and services most prone to corruption
		Tender and award processes —————	— CompraNet*
		- Public procurement complaints registers —	— Integral System of Nonconformities (SIINC)
		- Meeting records -	<ul> <li>Record of the minutes on the Celebration of Meetings and Visits in the hiring procedures (Protocol of Action in the matter of public procurement)</li> </ul>
		Register of government projects —	— Contractual State of the Public Work
		- Public officials register ————————————————————————————————————	Register of Public Servants of the Federal Government (RUSP)
			RHNet
		Government grants —	<ul> <li>Cash Donation Information System Granted by the Federation (donations and budget)</li> </ul>
	SHCP —	- Spending	— Accounting and Budget System (SICOP)

 $<sup>{}^*\</sup>text{May}$  repeat itself in other clusters as it corresponds to more than one of the Guide's 30 Datasets.

Cluster	Public Institution	Related Dataset from the Guide	System/Register/Database
ı	AMEXCID -	International aid and financing	National Registry of Cooperation for Development (RENCID)
	CD —	Lobbying register	Lobbying Register
	CJF —	Court data	— Public Decisions and Resolutions
			— Novel or relevant criteria
	COFEMER -	Licenses, concessions and permits —	Federal Registry of Procedures and Services (RFTS)
	CS	Lobbying register —	Lobbying Register
	INDESOL -	Charity register —	Organizations of the Civil Society with Administrative Procedure and Resolutions
	INE —	Political parties finances	— National Registry of Suppliers (RNP)
	SE	Company register	Mexican business information system
	SFP	Records of changes in regulations	— Inventory of Internal Rules in Force
			Governmental Improvement Project Information System
		List of government contractors	- CompraNet*
		Audit data	— Annual Audit Program
			List of signatures of eligible external auditors
			— Comprehensive Audit System (SIA)
			Results of Audit and Control Actions
			Computing System of Social Comptrollership (SICS)
			— Results of Audit and Control Actions (UORCS)
		Interest declarations	— Declaranet*
		- Asset declarations -	
		List of government contractors	Sole Registry of Suppliers and Contractors (RUPC)
		- Tender and award processes	— Buying Units
		Register of government projects	Electronic Log of Public Works for the Federal Public Administration (BEOP)
		- Public officials register	Register of sanctioned public servants (RSPS)
	SHCP		Registry of Public Servants of the Federal Public Administration that intervene in public contracting procedures, the granting of licenses, permits, concessions and authorizations, as well as in the alienation of movable property of the federal public administration and in the assignment and issuance of opinions in subject of appraisals and income appraisal (RENIRESP)
		Government grants	Comprehensive Information System of Patterns and Government Programs (SIIPP-G)
		- Spending	— Timely Public Finance Statistics
		Impuestos —	Collection of tax revenues from the Federal Government
		Politically exposed people's list	<ul> <li>List of public positions that are considered nationally politically exposed persons</li> </ul>
		- Budgets	- Budget Transparency

<sup>\*</sup>May repeat itself in other clusters as it corresponds to more than one of the Guide's 30 Datasets.

Cluster	Public Institution	Related Dataset from the Guide	System/Register/Database
2	COFEMER -	Records of changes in regulations ———	Regulatory Draft sent by the agencies to COFEMER
	INDESOL -	— Charity Register————————————————————————————————————	Federal Register of Civil Society Organizations (OSC)
	INEGI -	— Company Register ————————————————————————————————————	National Statistical Directory of Economic Units (DENUE)
	PGR —	Court data	Criminal incidence of federal jurisdiction
			<ul> <li>Statistics of the General Direction of Control of Criminal Proceedings and Appeals in the Matter of Federal Crimes (DGCPPAMDF)</li> </ul>
	SAT —	Charity register ————————————————————————————————————	— Directory of Authorized Grantees
	SFP —	Debarred or sanctioned contractors ———	Directory of sanctioned Suppliers and Contractors (DPCS)
		— Audit Data —————————————————————————————————	— Results of Appointment of External Auditors
	SHCP	- PPPs	Registration of Pure Public-Private Partnerships
			Registration of Projects of Public-Private Partnerships
			Registration of the investment work portfolio
		Spending —	Assignment and execution of the budget of expenses of the federation
			— Federal entities projects
			Federalized spending contracts
	SSA —	Charity register ————————————————————————————————————	— Amounts granted in joint venture projects with civil society organizations
	SE —	— International aid and financing ————————————————————————————————————	List of the Registry of Foreign Companies registered in the National Registry of Foreign Investments that perform acts of commerce in Mexico
	SFP —	Lobbying register	— Open Contracting