open
data
charter

Open Up Guide:
Using Open Data to Combat Corruption

TRANSPARENCIA MEXICANA

Open Contracting Partnership

OpenData

This practical resource identifies priority datasets, open standards and open data use-cases that governments, civil society and other stakeholders can focus on to tackle corruption at all levels and to respond to increasingly complex corruption networks.

Contents

Contents 2

Acknowledgements 3

Background: Charter Open Up Guides 4
The International Open Data Charter 4
Open Up Guide 4
Development process 4

Summary 5
Case studies 5
What needs to be in place 6
What's in the report 7
Next steps 7

Section 1: Analytical framework 8
Introduction: Meeting the challenge of corruption 9
Background: Developing the Anti-Corruption Open Up Guide 12
Strategy: Linking open data to anti-corruption strategies 13
–The Public Policy Cycle 13
–The Anti-Corruption Cycle 14
–Increasing responsiveness 15
–The role of open data in the anti-corruption policy cycle 17

Section 2: Anti-Corruption Open Data 20
Overview: Data against corruption networks 21
–Registers 22
–Transactional Data 22
–Public disclosures 23
Summary: priority datasets 24
–Availability challenges 25
Foundations of a solid anti-corruption data infrastructure 26

Section 3: Making use of open data 29
From data gathering, to data use 30
Prevention 31
–In the private sector 31
–In the public sector 32
Detection 33
Investigation 34
Enforcement 35

Section 4: Conclusions 36
Conclusions 37

Appendix 1: Datasets in detail 38

Acknowledgements

The Anti-Corruption Open Up Guide was supported by the Open Data for Development (OD4D) program, a partnership funded by Canada's International Development Research Centre (IDRC), the World Bank, United Kingdom's Department for International Development (DFID), and Global Affairs Canada (GAC).

Lead author: Rafa Garcia Aceves

Contributors: Tim Davies, Maria Zilli, Zainab Harrak, Eduardo Bohórquez, Piper Hart, Karolis Granickas, Lindsey Marchessault, Ania Calderón, Enrique Zapata, Fabrizio Scrollini

Series Editor (2016): Tim Davies

With thanks to everyone who shared participated in workshops, shared comments on the drafts, and input into this first version of the Anti-Corruption Open Up Guide.

An initiative by the:

open data charter

A collaborative product of:

Open Contracting Partnership

OpenData for development network

Coordinated by:

TRANSPARENCIA MEXICANA

Background: Open Up Guides

The Open Data Charter

The **Open Data Charter** is a movement of governments and organisations committed to achieving a world where government information is used to empower people in order to build more prosperous, equitable and just societies. The Charter was founded in 2015 and is based on six principles for how governments should be publishing information. They are:

1. Open by Default
2. Timely and Comprehensive
3. Accessible and Useable
4. Comparable and Interoperable
5. For Improved Governance and Citizen Engagement, and
6. For Inclusive Development and Innovation.

More than 70 governments and organisations have joined the movement. The Charter's goal is to embed the culture and practice of openness in governments in ways that are resilient to political change and driven by user demand.

Open Up Guide

A Charter Open Up Guide is a curated resource, providing insights and guidance on how to make open data available, and to support the use of open data, in a particular sector.

The guides are designed to support strategic action and data interoperability. They identify key datasets, common standards and practices that charter supporters can implement.

The guides are peer-reviewed and are developed through collaboration that engages both data publishers and data users. The guides are designed to be globally relevant: catering for the differing needs and capabilities of Charter supporters.

Development process

Development of the Anti-Corruption Open Up Guide began at the Open Government Partnership 2015 Summit in Mexico City. An initial consultation set the shape and focus of the guide, and in January 2016, Transparencia Mexicana were commissioned to lead authorship of the first draft. Workshops in London, Washington and Lagos informed the drafting process, with an beta version presented at the International Open Data Conference in Madrid in October 2016.

Overview

Corruption has a devastating impact on the lives of people around the world. When money that should be spent on schools, hospitals and other government services ends up in the hands of dishonest officials, everyone suffers.

A broad understanding of corruption recognises that it is not just about isolated acts between two different agents: the one who offers a bribe, and the one who receives it. Instead, corruption is a complex crime. It is driven by networks of officials, professional intermediaries and companies. So in order to tackle corruption effectively, you need to understand and dismantle these networks. This requires information and the ability to spot patterns.

Many of the activities of a corruption network, and many of the individuals and organisations involved, leave their mark on government held datasets. Paradoxically, corruption schemes frequently rely upon the law to secure ownership of companies, land and assets used to launder their proceeds. Public contracts, spending and other transactions are all recorded in government ledgers. And existing policies may call for asset disclosures and interest registers to be maintained. If all this information remains in silos, identifying, tracking and confronting corruption networks remains a laborious task.

That's where open data can step in. Open data is information anyone can access, use and share. It should be made available with the technical and legal characteristics necessary for it to be freely used, reused, and redistributed by anyone, anytime, anywhere. Publishing government information in this way has the potential to allow government officials, journalists and citizens to follow financial flows, understand who's providing government services and to spot suspect behaviour. There is political momentum behind this, and in 2015 the G20 agreed a common approach for how to use open data to combat corruption[1].

Case studies

Examples of where open data has been used to fight corruption include:

- In 2015, Global Witness published an investigation uncovering the powerful military, government and narcotics actors benefiting from Myanmar's jade wealth, and the way in which they are using a web of anonymous companies to hide their gains at the expense of the rest of the population. This was in part based on data made accessible as open data through OpenCorporates.com.
- Ukraine's public procurement system was once notorious for corruption and inefficiency. **Since launching ProZorro,** the country's open source, open data e-procurement system the government has saved 14% on its planned spending (more than 300 million Euros) and seen a 50% increase in companies bidding for contracts - helping build business and citizen trust in the government process.

[1] www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf

- In Latin America, the PODER network have built the 'QuiénEsQuién.Wiki' platform thato combines procurement information and company ownership information to support journalistic investigations.
- Open Contracting Partnership have identified 150 suspicious behavior indicators, or "red flags", that governments or civil society can use to identify potential corruption in procurement.

What needs to be in place

The initial premise that open data alone would lead to a radical change in outcome for governments and citizens has proved overly-optimistic. As a recent Transparency International/Web Foundation report found, countries are often failing to do the first step of following through on their commitments to put the right policies in place, let alone ensure that they are delivering results[2]. Where information is published, there are some important things that need to be in place if open data is going to have an impact on corruption. They are:

- The existence of information in digital, shareable formats. Lots of government processes use paper forms, or generate PDFs, which can be difficult and costly to digitise for computers to use.
- Information needs to be published as open data in line with the six principles of the Open Data Charter. The Charter sets out the best way for governments to publish data - it should be open by default, interoperable and timely.
- The right data needs to be published. Given that corruption happens through often complicated networks, you need multiple datasets to build a picture of what is going on. The guide includes a list of 30 different datasets that can be used to fight corruption.
- Datasets need to be able to talk to each other in order to build a picture of what's going on. The list of datasets in this guide includes information on what key features the datasets have to have in order to be interoperable. The Guide helps governments understand what datasets they should prioritise for collection and release, and which data standards are out there to use.
- Citizens, journalists and members of civil society have the skills and legal protection to analyse the data that's published, share findings publically and seek a response to their findings. The direct users of raw data are likely to be a small number of engaged specialist organisations and individuals, who can act as intermediaries between data sets and the broader public.
- Data publication and analysis needs to be accompanied by other investigative methods, such as interviewing sources. Insights from a dataset is often only the first step in combatting corruption.
- There need to be structures in place so that governments and law enforcement respond to indications of potential corrupt activity.

[2] www.transparency.org/whatwedo/publication/connecting_the_dots_building_the_case_for_open_data_to_fight_corruption

What's in the report

This Open Up Guide report is intended to be part of a growing body of work using open data to fight corruption. It includes:

- Use cases and methodologies. A series of case studies highlight existing and future approaches to the use of open data.
- 30 priority datasets and the key attributes needed so that they can talk to each other[3]. To address corruption networks it is particularly important that connections can be established and followed across data sets, national borders and different sectors.
- Data standards. Standards describe what should be published, and the technical details of how it should be made available. The report includes some of the relevant standards for anti-corruption work, and highlights the areas where there are currently no standards.

Next steps

The Open Data Charter will work with a number of governments to test the hypothesis behind the Guide that open data can help tackle corruption. The Charter team is in early conversations with the Government of Mexico about developing a methodology to road test the guide and ground this work to address real life challenges.

This Guide is a living document and we're planning to update it. Please feel free to add any comments to this Google doc, or get in touch with the Charter team directly at: **info@opendatacharter.net.**

[3] See the full list of datasets here: https://airtable.com/shrHY9KFJ5bircwvx.

Section 1: Analytical framework

Transparency plays an important role in challenging corruption, but it is important to go beyond the idea that disclosing data directly equals reduced corruption. Open data can only play an effective role in dismantling corruption networks, if governments secure its availability and interoperability, and all stakeholders put open data to use as a practical tool.

Introduction: Meeting the challenge of corruption

Where's our starting point?

There are many definitions of **corruption** in use around the world, some more detailed or precise than others. These definitions vary across different countries and legal systems[4]. As a global reference, and as a starting point for this resource we use the **Transparency International** definition of corruption as "the abuse of entrusted power for private gain"[5]. This short but widely accepted definition contrasts with the broad array of forms that an act of corruption can take: from bribes to local officials, to sophisticated financial schemes for transfer mispricing or tax evasion.

Box 1. Is it possible to classify corruption?

Different scholars and practitioners have tried to propose analytical categories to understand and investigate corruption. Depending on the angle and objectives of the researcher, categories of corruption vary. Transparency International have identified at least three broad categories[6] that allow to group and analyze certain acts of corruption:

1. ***Grand corruption*** *is the abuse of high-level power that benefits the few at the expense of the many, and causes serious and widespread harm to individuals and society. Such acts distort policies or the central functioning of the state, enabling leaders to benefit at the expense of the public good. Transparency International has proposed an specific legal definition that can be consulted here.*

2. ***Petty corruption*** *refers to everyday abuse of entrusted power by low- and mid-level public officials in their interactions with ordinary citizens, who often are trying to access basic goods or services in places like hospitals, schools, police departments and other agencies.*

3. ***Political corruption*** *is a manipulation of policies, institutions and rules of procedure in the allocation of resources and financing by political decision makers, who abuse their position to sustain their power, status and wealth.*

Source: Transparency International

Frequently, descriptions of corruption are over-simplified, reducing them to isolated acts between two different agents: the one who offers a bribe, and the one who receives it. In other words, as the often repeated quote puts it: "*it takes two to tango*". Yet in reality, corruption is a complex crime. Many modern cases of corruption are the result of a series of actions and agreements to subvert government processes, influence or manipulate policies and oversight mechanisms, extract rents, and hide the proceeds. Committing an act of corruption requires the complicity, and secrecy, of a group of individuals and organizations. In other words, a **corruption network**.

As institutional and legal frameworks adjust and improve to reduce spaces for corruption, these networks also adjust and increase their sophistication and complexity. Analytical approaches to understand the changing dynamics of corruption networks are diverse. Often, only some parts of

[4] Different national legal systems and international agreements each use their own definitions of corruption, or set out a specific series of acts or forms of conduct that are considered as corruption. Penal Codes reviewed: Canada, Czech Republic, Finland, France, Greece, Hungary, Italy, Japan, Luxembourg, Korea, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, United Kingdom, United States and Turkey.

[5] Transparency International. Available at: https://www.transparency.org/what-is-corruption#define

[6] Idem.

the network are 'visible', or in the reach of particular authorities. As a result, a wide range of conventions, laws, policies, or mechanisms intended to reduce risks and impact of corruption schemes have been designed and implemented across the world. Some of them with better impact than others. **Open data** could play a key role, enhancing the effectiveness of this set of anti-corruption tools.

In 2014, the G20 Anti-Corruption Working Group (ACWG) recognised the importance of **open data as an emerging resource that could be used to address corruption**, highlighting its potential to follow financial flows, open up public contracting and procurement, change incentives in corruption-prone environments, and to enable cross-sector collaboration[7]. From extractives and aid, to procurement and political finance, initiatives have turned to open data, as a powerful tool to identify, and ultimately dismantle, corruption networks.

Box 2. Corruption: a global agenda

In the last two decades, corruption has been a headline issues in the global agenda. Government leaders, civil society organizations, private sector leaders and international agencies have highlighted the effects of corruption on global stability and inclusive development. Several global initiatives have called for robust and sound legal frameworks that deals with corruption issues, such as bribery, as well as spaces for coordination among governments and institutions. As a response, the global community has reached specific agreements and commitments to fight and control corruption. Some of them are:

- The **OECD Anti-Bribery Convention** is a global instrument that makes "foreign bribery a crime for which individuals and enterprises are responsible" (Exporting Corruption Report, Transparency International)[8]. It establishes legally binding obligations with the purpose of preventing and sanctioning bribery of foreign public officials in international business transactions. Since its publication in 1997, 41 countries have adopted and ratified it.

- The **United Nations Convention Against Corruption (UNCAC)** was adopted by a UN General Assembly resolution of October, 31st, 2003 and on December, 9th 2003, 114 countries signed the Convention in Merida, Mexico. The treaty focuses on combating money laundering and promoting integrity within governments through prevention, criminalization, international cooperation and asset recovery.

- The **G20 Anti-Corruption Working Group (G20 ACWG)** was established in Toronto in 2010 as a response to the impact of corruption and bribery on international economic growth. The agenda of the group includes beneficial ownership transparency, bribery, corruption in high-risk sectors, public sector transparency and integrity, international cooperation and private sector transparency and integrity. The G20 ACWG has also adopted a set of principles on the use of open data against corruption, although a February 2017 report by Transparency International and the Web Foundation found that a number of G20 countries were failing to live up to these commitments[9].

- The **UK Anti-Corruption Summit** was an international event hosted last May 12th, 2016 by the UK Government in London, to "galvanise a global response to tackle corruption"[10]. 40 countries attended the Summit where governments signed up to a set of general principles and signed a global declaration against corruption. Some governments made specific country commitments.

These initiatives show the different angles for addressing corruption issues from a global perspective. Their effectiveness is based on both implementation by member countries, and coordination among

[7] G20. Available at: http://www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf
[8] Transparency International. Available at: http://www.transparency.org/exporting_corruption
[9] TI/Web Foundation. Available at: https://www.transparency.org/whatwedo/publication/connecting_the_dots_building_the_case_for_open_data_to_fight_corruption.
[10] United Kingdom Government website. Available at: https://www.gov.uk/government/topical-events/anti-corruption-summit-london-2016/about

different actors.

Many of the activities of a corruption network, and many of the individuals and organisations involved, leave their mark on government held datasets. Paradoxically, corruption schemes frequently rely upon the law to secure ownership of companies, land and assets used to launder their proceeds: although they seek secrecy by using shell companies and complex ownership networks. Public contracts, spending and other transactions are all recorded in government ledgers. And existing policies may call for asset disclosures and interest registers to be maintained.

Whilst all this information remains in silos, identifying, tracking and confronting corruption networks remains a laborious task. If more data is made available and interoperable (increasing the ability to find connections between datasets), it should be possible spot suspicious patterns and so strengthen anti-corruption initiatives implemented either by governments or civil society organizations.

Box 3. What is Open Government Data?

During the past years, the **open data** agenda has grown significantly. Facilitated by technology and digital media, what started being a technical and specialized discussion has now become a global movement. The International Open Data Charter (IODC) defines open data as "**digital data that is made available with the technical and legal characteristics necessary for it to be freely used, reused, and redistributed by anyone, anytime, anywhere**"[11]. Many governments around the world have embraced this agenda, recognizing the value of open data *"to advance collaboration around key social challenges, provide effective public oversight of government activities, and support innovation, sustainable economic development, and the creation and expansion of effective, efficient public policies and programs"*[12].

For more information visit the International Open Data Charter "Resource Center" at:
www.opendatacharter.net/resource-centre/

However, open data is not a magic bullet. The increasing release of datasets on governmental budgets, contracts and audits, politicians and public servants' assets, central to the traditional transparency and accountability agenda, does not automatically translate into effective anti-corruption policies, mechanisms or tools. Even when it is possible to find some cases where open data has been a valuable tool for anticorruption, there is still a general perception that more can be done. Whilst recognizing the importance of transparency to inhibit corruption, **it is key to go beyond the idea that disclosing data directly equals corruption reduction**[13].

The release of open data must allow the development of specific tools that are able to activate authorities and institutions with anti-corruption responsibilities and capabilities. It must also allow citizens and journalists to connect and track information across different datasets, fostering more effective civic anti-corruption initiatives. Understanding the dynamics of anti-corruption data, structuring and publishing it in open formats and improving its quality are initial tasks to enable the use of technology and the development of innovative applications to fight corruption networks. Further actions, such as investing in capacity-building for data analysis

[11] International Open Data Charter. Available at: http://opendatacharter.net/principles/
[12] Ídem
[13] Eduardo Bohórquez, Irasema Guzmán, German Petersen, "Factofilia: Más transparencia no es igual a menos corrupción", Este País: tendencias y opiniones, (01/10/2015), Available at: http://www.estepais.com/articulo.php?id=291&t=factofilia-nbspmas-transparencia-menos-corrupcion.

of government agencies, prosecutors, journalists and civic watchdog organisations, will also be needed.

By acknowledging the complexity of this discussion, the **Anti-Corruption Open Up Guide** is intended to be a starting point for a structured dialogue among different actors –governments, civil society organizations, international organizations, investigative journalists, business associations, small and medium enterprises, entrepreneurs– on the use of open data for preventing, detecting, investigating and sanctioning corruption. It is not intended to be an exhaustive tool but rather the basis for even more focussed work around specific corruption challenges.

The Anti-Corruption Open Up Guide identifies:

- **Use cases and methodologies**. Increased disclosure alone does not equal reduced corruption. Open data needs to be used to be effective and in combination with other tools and sources of information. A series of case studies highlight existing and future approaches to the use of open data, as well as relevant contextual factors.

- **Priority datasets and their attributes**. To address corruption networks it is particularly important that connections can be established and followed across information systems, national borders and different sectors. This calls for shared prioritisation of key datasets.

- **Data standards.** Standards describe what should be published, and the technical details of how it should be made available. They play a key role in bringing about data interoperability.

Background: Developing the Anti-Corruption Open Up Guide

Box 4. Developing the Anti-Corruption Open Up GUide

The Anti-Corruption Open Up Guide was developed through four core activity tracks. These were carried out simultaneously and are interconnected, as information gathered in the four of them contributed with ideas, experiences and databases to shape this resource. This work was complemented by a range of workshops, and engagement around the 2016 Anti-Corruption Summit in London.

[Track 1] Building a shared language: framework discussion	[Track 2] An initial input: key anti-corruption data
Based on the review of global initiatives, legal frameworks, academic papers and interviews with anti-corruption practitioners, an analysis of how corruption works and links to public policies was developed. It responds to the multiple points of entry of the anti-corruption discussion to set out a common framework for understanding data use, and for prioritising datasets. The resulting framework is used to guide the user throughout the resource.	Through an online consultation, a series of interviews with government officials, investigative journalists, researchers and civil society organizations and a review of the open data portals from G20 countries, a full range of datasets with the potential to play a role in anti-corruption work were identified. These were organized according to their anti-corruption value and their data characteristics as well as tagged to thematic areas where they can be used. The final priority list was based on assessment against use-cases.

[Track 3] Acknowledging anti-corruption efforts using open data: relevant experiences	[Track 4] Frontiers, challenges and ideas: potential uses of open data for anti-corruption
After a series of online and offline consultations, cases where data or open data had been used for anti-corruption purposes were identified and reviewed. Selected cases were analyzed in-depth, reaching and interviewing their implementers. These cases were a source both of relevant datasets and examples on how open data can be used to prevent, detect, investigate or sanction corruption. Most of the cases identified were related to public procurement.	Responding to innovative anti-corruption ideas shared through the online consultations, the report has identified gaps in current data availability. A review was carried out of the corruption problems experienced by a number of different countries that might be addressed with open data in future. These sources of information led to the development of future potential uses cases of open data for anti-corruption.

Strategy: Linking open data to anti-corruption strategies

What should be taken into account?

To understand the role open data can play within anti-corruption strategies, as well as to identify the relevant data at the right time, it is necessary to acknowledge the existence of **two different but simultaneous policy cycles**. The first one is the public policy cycle, which can run from the moment a problem or a need is defined, through the planning, resource allocation, implementation and evaluation of the policy (*see table 1*). The second one is the anti-corruption cycle, identifying different stages involved in tackling corruption (*see table 2*). Both cycles are intrinsically interconnected since any public policy can be affected by corruption.

Ideally, every public policy should be accompanied by clear anti-corruption policies and mechanisms. In considering how open data can enhance these two policy cycles, it will be important to consider the data, and methods for its use, that are most relevant at each stage of the policy in question and the corruption risks identified.

The Public Policy Cycle

There are different versions of the **public-policy cycle** in the literature. Taking different models into account, we divide it broadly into five stages: problem identification, policy development, resource allocation, implementation and evaluation. Different policy areas will face different corruption risks at each stage. In the table below we present a number of examples of potential risks and strategies that respond to them.

Table 1. The public policy cycle

Cycle stage	Example Corruption risks	Example Anti-corruption strategies

Problem identification: when an issue becomes the focus of resources and new policy.	Defining unnecessary needs, or defining needs in a way that could only be met by a particular actor.	Mandatory asset declarations, registers of interests, lobbying registers and political finance transparency.
Policy development: when objectives, methodologies and rules for delivering the policy are set.	Designing policies to benefit particular firms. Adopting approaches that minimise open competition, or increase discretionary resource allocation.	Standard procurement processes and rules. Independent oversight of decisions.
Resource allocation: when budgets are set, and decisions about their distribution are made (e.g. through contracts).	Deliberate overestimation and padding of budgets. Awarding contracts to phantom firms, or selecting firms in return for kickbacks.	Budget and spend transparency. Use of e-procurement systems and civic oversight mechanisms (for example, Social Witness).
Implementation: when the policy is being executed.	Delivering substandard goods and services. Biased choice of subcontractors. Modifying plans or contracts to increase payments. Failing to enforce penalty clauses.	Whistleblowing policies. Social Audits. Mechanisms for feedback public redress.
Evaluation: when the effects and integrity of the policy are evaluated.	Covering up failures of delivery.	Government audits. Blacklisting.

The Anti-Corruption Open Up Guide

The Anti-Corruption Cycle

The **anti-corruption cycle**[14] can be broadly divided into four stages: prevention, detection, investigation and sanction. The use of these four stages gives room for analyzing different policies despite governmental differences, which is essential for keeping a global perspective. Some countries may have more developed institutional frameworks in some stages of the anti-corruption cycle. For example, the UK has strong prevention processes, for instance, whereas emphasis in the United States or Brazil is often in terms of enforcement. However, all countries share these four stages of anti-corruption action. Even when the organizational arrangements of governmental institutions are different, this cycle allows to identify the potential users of this resource in each stage based on their anti-corruption functions or activities.

Table 2. The Anti-Corruption Cycle

Stages of the anti-corruption cycle	Definition	Potential actors involved
Prevention	Actions, mechanisms and tools that reduce risks or increase the barriers and costs of corruption within a policy or procedure.	Policy makers (government agencies), parliamentarians, regulatory bodies, civil society organizations (CSO's), international

[14] Consultation for the Anti-Corruption Sector Guide. Available at: The Consultation of the Anti-Corruption Package is Launched at G20 Summit https://goo.gl/GLtvFt.

		agencies.
Integrity breach: commitment of an act of corruption		
Detection	Actions, mechanisms and tools that identify an illicit behaviour as a result of corruption.	Policy implementers (government agencies and internal control bodies), CSO's, investigative journalists, oversight institutions (auditors, comptrollers, parliament).
Investigation	Actions, mechanisms and tools intended to expose[15] and compile information about the illicit behaviour detected and the parties involved.	CSO's, investigative journalists, oversight institutions (auditors, comptrollers, parliament), national and foreign law enforcement and prosecuting institutions.
Sanction	Actions, mechanisms and tools intended to prosecute and penalize corruption crimes, as well as to recover illegal rents stolen or generated by a corruption crime.	Oversight institutions (auditors, comptrollers), national and foreign prosecuting institutions, judiciary, asset recovery agencies.
The Anti-Corruption Open Up Guide		

Increasing responsiveness

Although in theory anti-corruption and policy cycles should coexist, both oriented towards the effective use of public resources to guarantee the rights of citizens or deliver public goods or services, in practice there can be tensions between them. Whilst public policy aims to deliver solutions to concrete problems in the most effective and efficient way possible, anti-corruption strategies place emphasis on compliance with rules and procedures to ensure the integrity of the policy process.

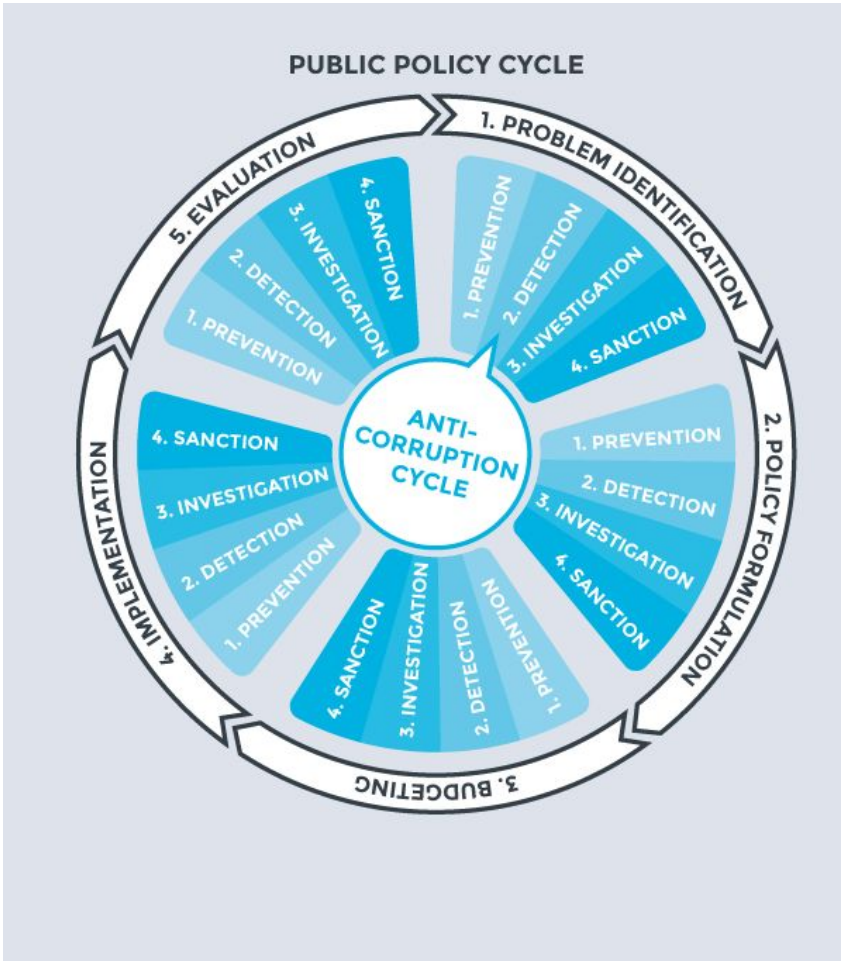
A specific public policy, considered against the goals that were set out for it, may be efficient and effective, and yet at the same time be affected by vested interests or even by corruption. For example, when analyzing a corruption case, an investigative journalist may focus on the budgetary stage of a public policy, revealing information that is crucial for illustrating misconduct of a public official, but not essential for securing the effectiveness of the public policy in which the integrity breach occurred. Alternatively, an anti-corruption policy may reduce the risk of corruption or prosecute a case effectively, while suspending the expected outcomes of a concrete public policy or delaying its implementation. A feeling that anti-corruption requirements can create additional bureaucracy and delays to processes can create challenges when governments are facing demands to become more responsive.

It is crucial then to explore how open data can be introduced and integrated into the policy processes in ways that protect and promote the responsiveness of policy to citizen need, whilst also increasing responsiveness to corruption risks and incidences. In line with it, the **Anti-Corruption Open Up Guide** will mainly focus on the ways open data can strengthen the effectiveness of the anti-corruption policy cycle. In other words, it will review how governments,

[15] Investigation of corruption cases may imply certain degree of secrecy according to the type of crime committed. Certain data should be held temporarily to avoid affecting investigation and enforcement.

law enforcement, civil society or journalists can use open data to become more responsive to corruption risks and cases.

Image 1. The link between the public policy cycle and the anti-corruption cycle



During any stage of a public policy, one or more corruption activities can happen at the same time.

The role of open data in the anti-corruption policy cycle

Data generated by government is a key resource to prevent, detect, investigate or sanction corruption. Almost every act of governments today can leave a trace in documents and data, and governments impose legal obligations on many of the other individuals and organisations who may be involved in corruption networks, requiring them to register or disclose key information. These records can be organized and stored in many different ways. With increasing government use of technology, records are frequently generated and stored through the use of **information systems**, organized and structured in state held **databases**. Some of the information systems governments used are known by the public, such as public procurement systems. But there are many more out of sight of the public that are used regularly, for example, to keep track of budget expenditure or monitor security in cities.

When data is **open by default** the possibilities of involving a higher number of different actors in anti-corruption efforts increase, allowing them to also form **anti-corruption task forces**. Depending on the available data, who is using it and the anti-corruption objective set, a wide range of uses can be designed (*see table 3*). Moreover, when the available data is structured in open formats, is free, accessible across borders and possible to archive, the probabilities of being effectively used for anti-corruption purposes increase.

Table 3. Main uses of data along the anti-corruption cycle

Stage of the anti-corruption cycle	Potential data user	Main data use	Anti-corruption objective
Prevention	Policy makers, regulatory bodies, oversight institutions (auditors, comptrollers) and CSO's	**Release** and **publication** of open data	Increase **transparency**, openness and oversight
		Identify potential corruption **risks**	**Mitigate** potential corruption risks and **amend** or **improve** public policies or regulation
Integrity breach: commitment of an act of corruption			
Detection	Policy implementers, CSO's, investigative journalists, oversight institutions (auditors, comptrollers, parliament)	Generate **alerts** about corruption acts. Identify corruption **networks**	**Activate** investigative anti-corruption institutions and mechanisms
Investigation	CSO's, investigative journalists	Expose a **corruption cases** increasing public understanding of a corruption scheme	Increase **social demand** for formal investigation, sanction, legal or policy reform
	Oversight institutions (auditors, comptrollers, parliament), prosecuting institutions.	Gather **evidence** about a corruption network, its arrangements and schemes.	Strengthen **prosecution** process and activation of administrative or penal prosecuting channels
Enforcement	Integration of **legal records** and court documents.	Achieve effective **sanctioning** and asset recovery.	Punishment of corruption and the recovery of assets to be spent on government priorities

The Anti-Corruption Open Up Guide

Identifying which government databases can be related to corruption is not easy. As forms of corruption vary across countries and legal frameworks, also government systems and databases vary. Take as an example, budgetary datasets. Accounting standards set by law, or by the Ministry of Finance of a country, affect how budget data is structured. Also, disclosure obligations for interests or company registers vary widely across countries. Taking into account that corruption operates through **complex networks**, which leave data footprints throughout

different databases, it must be acknowledged that the greatest power of data comes when users are able to combine datasets. The adherence to global data standards is a valuable tool to reduce variations across datasets and to allow cross-references between databases, cross-country comparisons and more complex anti-corruption investigations.

Section 2: *Anti-Corruption Open Data*

*A solid **anti-corruption data infrastructure** needs a range of different datasets, published to a high quality and in ways that allow connections to be made across between them. The adherence to **open data standards** contributes in securing that a larger number of users can benefit from the data available.*

Core data for setting an anti-corruption data infrastructure

Acknowledging the way corruption works, we have identified datasets relating to each of the **core elements of a corruption network**: a group of **individuals and organizations**, organized through a series of agreements and schemes – in some cases violating laws and **government procedures**– to extract certain **rent** from the public or obtain an undue benefit for a **private gain** (*see table 4*).

These datasets form a basic core that countries should strive to make available and interoperable. The approach is of course a general one as what data is available, and in what format, will vary from country to country, and case to case. Nor is this list definitive: there are many other datasets that can be relevant to specific anti-corruption efforts. However, together these datasets form the basis of a solid **anti-corruption data infrastructure**.

Table 4. Classifying anti-corruption data

Core element of a corruption network	Description of the related data to the core element	Examples of datasets
Individuals and organisations	Refers to any dataset containing records and information on entities (individuals or organisations) that can be potentially involved in a corruption scheme. Datasets under this category should provide information about the nature and characteristics of any entity, as well as its connections with others.	– Lobbying registers – Company registers – Interests registers – Politically exposed people registers – Advisory boards – Government contractors – Public servants directories – Charity registers
Public-related resources	Refers to any dataset containing records and information on the resources which belong to governments or are intended for public purposes and that could be involved in a corruption scheme. Datasets under this category should provide information about the status and transactions related to those resources.	– Budgetary datasets – Government spending – Contracts – Public-private partnerships – Political financing – Licenses and permits – Grants and scholarships – Auditing datasets – International aid, funding & technical assistance register
Regulations, rules and government procedures	Refers to any dataset containing records and information on the channels used, avoided or violated to commit an act of corruption by a corruption network. Datasets under this category should provide information about the procedures, events and legal acts potentially linked to corruption schemes.	– Voting records – Meeting records – Court records – Campaign promises
Rent extractions	Refers to any dataset containing records and information on the use of public	– Assets declarations/registers – Cadastre (including public land)

	resources that were potentially extracted as a result of a corruption scheme. Datasets under this category should provide information about the income sources and ownership of the assets owned by members of a corruption network.	– Land and property registers – Tax databases – Customs data

The Anti-Corruption Open Up Guide

These datasets take many different forms. Data may be drawn from **public registers** created to serve broad public functions, or developed with specific transparency and anti-corruption goals in mind. It may be **transactional data** generated during the daily operations of government, and released in as close to real-time as possible. Or it may be drawn from **public disclosures** mandated by law or policy.

Registers

Governments manage many different registers: from company registers, and land ownership registers, to lists of registered lobbyists, or lists of public servants.

The UK Government Digital Service (GDS) describe a register as *"...an authoritative list of information you can trust"*[16]. This is an ideal. Every effort should be made to ensure government registers are authoritative. GDS have developed principles for public registers, and an open source software stack that provides open APIs for access to 'living registers'[17].

However, sometimes government registers are not kept up to date, or they are maintained in non-interoperable and error-prone ways. This can lead to third-parties maintaining their own open data registers based on aggregating together and checking on the quality of government provided data.

Box 5. Registers: Every Politician

EveryPolitician.org is an independently maintained datasets with the goal of providing *"data about every national legislature in the world, freely available for you to use"*. Using the Popolo standard to manage data, its dataset is populated by a mix of 'screen-scraping' official resources, and crowdsourcing information.

If governments provide official registers of political figures, then the EveryPolitician Bot can more easily keep the platform up to date.

The Anti-Corruption Open Up Guide

Transactional Data

Every day hundreds of land deals take place; thousands of government tenders are issued, and contracts signed; and millions of payments may be made to and from government.

Hidden within these transactions may be red-flags for corruption, or information that, when

[16] https://gds.blog.gov.uk/2015/09/01/registers-authoritative-lists-you-can-trust/
[17] https://gds.blog.gov.uk/2015/10/13/the-characteristics-of-a-register/

linked with information from a register, could show illicit benefits received by a government official.

Transaction data can be made available in real-time through APIs, or provided periodically in bulk downloads. Timeliness and disaggregation can be an important factor in the use of transactional data, but care must also be taken to respect privacy.

Box 6. Transactional data: Brazil's transparency portal

Brazil's Transparency Portal provides detailed data on five key categories of transaction:

1. *Direct spending by federal government agencies through contracts and tender processes;*
2. *All financial transfers to states, municipalities and the federal district;*
3. *Financial transfers to social program benefactors;*
4. *Administrative spending, including staff salaries, staff travel expenses and per diems and office expenditures; and*
5. *Information on all government official credit card spending*".

Some transactional information is updated on a nightly basis. The portal has over 900,000 unique visitors per month.

Source: http://odimpact.org/case-brazils-open-budget-transparency-portal.html. The Anti-Corruption Open Up Guide

Public disclosures

Transparency policies often create an obligation on public bodies, public figures or private entities to disclosure information. For example, disclosing a record of meetings between lobbyists and officials, or publicly posting voting records. Sometimes this information is recorded in registers, but often the obligation is worded so that bodies post their own disclosures on local notice boards, websites or in gazettes.

Frequently such disclosures are made in non-standard formats, in word processed documents, making it difficult to join up this information to other datasets. If standard formats were used, and data was more easily discoverable, the anti-corruption value of these disclosures could be increased.

Summary: priority datasets

Table 4. Summary of priority datasets for building an anti-corruption data infrastructure

Dataset	Anticorruption category	Type of dataset
Interest declarations	Individuals and organizations	Register
Lobbying register	Individuals and organizations	Register
Company register	Individuals and organizations	Register
Charity register	Individuals and organizations	Register
Politically exposed people's list	Individuals and organizations	Register
Public officials register	Individuals and organizations	Register
List of government contractors	Individuals and organizations	Register
Corruption-sensitive postings	Individuals and organizations	Register
Council / advisory board members	Individuals and organizations	Register
Contracts register	Individuals and organizations	Register
Political parties finances	Public-related resources	Public disclosures
Budgets	Public-related resources	Public disclosures
Tender and award processes	Public-related resources	Transaction
Licenses, concessions and permits	Public-related resources	Transaction
PPPs	Public-related resources	Register
Spending	Public-related resources	Transaction
Government grants	Public-related resources	Transaction
International aid and financing	Public-related resources	Register
Audit data	Regulation, government procedures and records	Transaction
Voting records	Regulation, government procedures and records	Register
Court data	Regulation, government procedures and records	Transaction
Register of government projects	Regulation, government procedures and records	Register
Meeting records	Regulation, government procedures and records	Register
Records of changes in regulations	Regulation, government procedures and records	Register
Campaign promises	Regulation, government procedures and records	Register
Debarred or sanctioned contractors	Regulation, government procedures and records	Register
Public procurement complaints registers	Regulation, government procedures and records	Register
Land/Property register and	Rent Extraction	Register

cadastre (public land)		
Tax records	Rent Extraction	Register
Asset declarations	Rent Extraction	Register

For full data description see "Annex 1" or access: https://airtable.com/shrtE30MaSKb1sjko. The Anti-Corruption Open Up Guide

Availability challenges

In many countries, open data remains a very recent policy agenda. Since 2009, an increasing number of countries, regions and institutions have launched open data portals, yet few of the datasets crucial to combatting corruption are currently open by default. The 3rd Edition of the **Open Data Barometer** measures the availability of open data across 92 countries, and covers five key accountability datasets: corporate registers, government spending, land ownership, contracting and budget. On average, less than 10% of them were available as open data (see table 5).

Table 5. Selected statistics from the 3rd Edition of the Open Data Barometer

Dataset	Percentage as fully open data[18]	Percentage available online
Corporate registers	**1%**	**72%**
Least open dataset in the world with just Australia publishing it as open data and only for very top level data for free. Looking at all datasets available, it is only accessible online in a machine readable format and for free in just a dozen of countries. The absence of adequate open data on companies makes tracking beneficial ownership challenging and hampers efforts to tackle corruption.		
Government spending	**2%**	**4%**
Weakest dataset in the study. Even in the limited cases when it is available online, the data is usually not published at the transactional level. Only four countries – the USA, the UK, Japan and Brazil – publish spending data at the transactional level and from those only two, the UK and Brazil, release this information as open data. This makes it nearly impossible for government, citizens and civil society to tackle corruption		
Land ownership	**5%**	**46%**
Rarely available online, difficult to find when available and quite frequently behind paywalls.		
Contracting	**8%**	**82%**
Only 28% of the data available online is in machine-readable formats reducing practical accountability as this makes analysing the high volume of historical data very difficult.		
Budget	**18%**	**97%**
Comparatively one of the better datasets in terms of availability and openness. In 95% of countries where it was available it was regularly updated. In several instances it is even required by law to be updated and available, although not necessarily open.		

[18] Meeting the open data definition

Foundations of a solid anti-corruption data infrastructure

Joining up data and standards for anti-corruption

A solid anti-corruption data infrastructure can only be built when the relevant datasets can communicate with each other. The higher the number of connections, the better the chance of using the datasets to spot potential corruption red flags. Based on the priority datasets for building an anti-corruption data infrastructure (*see table 4*), a series of core data elements have been identified and have also been matched to available **open data standards**.

A data standard is a framework for how data should be collected and published, including how to describe individuals and organisations, how to register specific events or transaction and how to organize data to meet minimum quality requirements. Using a standardised approach means that different datasets can talk to each other. Moreover, the adherence to open data standards contributes in securing that a larger number of users can benefit from the data available.

It is desirable that both governments and civil society, review the existing availability of data and agree on a route map to disclose it as open data. At the same time, it is important to review how data is structured and assess if it needs to be restructured to meet open data standards.

Table 5. Summary of priority data standards for building an anti-corruption data infrastructure

Name of data standard	Description	Sponsor
Open Contracting Data Standard	Data guidance for disclosing public procurement data in open formats about contracting processes from planning to implementation stage. Extensions for other types of contracting such as public private partnerships and concessions are under development. More information: http://standard.open-contracting.org/	Open Contracting Partnership (CSO)
Fiscal Data Package	Schema for publishing and consuming fiscal data, especially data generated during the planning and execution of budgets. It supports data on expenditures and revenues[19]. More information: http://specs.frictionlessdata.io/fiscal-data-package/	Open Knowledge Foundation (CSO)
Popolo	Popolo is an initiative on open government data specifications. Its goal is to "define data interchange formats and data models so that organizations can spend less time transforming and modeling data and more time applying it to the problems they face". It allows standardization of data related to people,	

[19] Open Knowledge Foundation

	organizations, motions and voting, events, speeches, among others. More information: http://www.popoloproject.com/	
Global Beneficial Ownership Register	An open schema under development for collecting and publishing beneficial ownership data globally. It will enable users to register in a standardized way data about the ultimate beneficiary or owner of a certain good (such as land) or an organization or entity (such as companies) across different countries. More information: http://openownership.org/get-involved/	Open Ownership (Global Coalition)
Open Corporates Schema	Schema for publishing and consuming data on companies worldwide, including data on jurisdiction, incorporation date, shareholders and subsidiaries. It recently incorporated beneficial ownership data released by the UK Government. More information: https://github.com/openc/openc-schema	Open Corporates (Private firm)

Box 6. The G20 Open Data Portals: enablers of Anti-Corruption Data?

The G20 has recently pushed the open data agenda globally. Accounting for 85% of the gross world product (GWP), 80% of world trade and two-thirds of the world population[20], actions implemented by these countries can lead trends across the world. Taking this into account, open data portals from the G20 countries were reviewed to understand the ease for identifying anti-corruption related datasets.

To start, only 16 out of 20 members have an open data portal. China, South Africa, South Korea and Turkey have not yet launched a portal where open government datasets can be accessed and downloaded. In total, these open data portals[21] contain 593,220 datasets. The top three countries with more datasets available are Canada (41.3%), the United States (33.7%) and the United Kingdom (4.4%).

Based on this sample, a series of related-corruption words —in the portal's official language— were looked up through their own search engines. For example, when the words "Corruption" and "Anti-corruption" were key search words, a total of only 114 and 311 datasets were respectively found. This means that **only 0.05% of the available datasets is directly classified as a resource that could be used for anti-corruption purposes**. Saudi Arabia, Mexico, Germany, Brazil and Argentina yield 0 answers for both requests.

Although, these results are not conclusive regarding the existence of anti-corruption data, they are prove that better categorizations or search mechanisms are needed to access such data. As matter of fact, the number of data fixed categories goes from 9 up to 33, making difficult to find data on similar issues across countries. Also, 50% of the open data portals reviewed (Australia, Argentina, Brazil, France, Germany, Indonesia, Japan and the USA) offer users the possibility of tagging freely datasets, allowing to search for information outside the standard categories. Regardless of the approach to be chose by each country, it is clear that there is a great opportunity for G20 governments to make their open data portals enablers of anti-corruption strategies.

[20] OECD 2015 numbers
[21] The G20 Open Data portals were reviewed in april 2016

Section 3: *Making use of open data*

Open data can be an anti-corruption resource for governments, civil society, journalists and the private sector. The appropriate strategies for data use vary between sectors, and the stage of the anti-corruption cycle that is being addressed.

From data gathering, to data use

Unlocking resources for anti-corruption action

Many of the pioneers of data-driven anti-corruption work have not been able to draw upon proactively published open datasets. Instead, they have had to gather the data they need through Freedom of Information requests, scraping data from inaccessible websites, and, in some cases, working with leaked datasets. Where data has been available, it has often been low quality, requiring substantial investments of time and effort before it can be used - and limiting the extent to which tools from one country can be used in another.

in this section we detail a number of cases where different stakeholders are working with the datasets described in the last section - either directly from open data, or using data they have manually gathered. The more governments move towards proactive publication, the more use-cases like these can spread, and effort can go into data use, rather than data gathering.

Over the coming year we hope to revise this section with additional cases that demonstrate

direct use of open data - as governments deliver on their commitments to provide structured open data for anti-corruption.

Case study: The Panama Papers

The Panama Papers are an unprecedented leak of 11.5 million files from the database of the world's fourth biggest offshore law firm, Mossack Fonseca. The leaked files reveal information on more than 214,000 offshore companies, connected to people in 200 countries and territories. The data includes emails, financial information, and corporate records that in some cases link world leaders and other prominent figures to illicit activity.

The International Consortium of Investigative Journalists worked with the leaked documents, and imported structured data extracted from them into a graph database, providing this to a network of 100s of investigative journalists. This made it possible to find leads in the dataset, and to follow up potential stories. The investigations and stories from analysis of this data have led to multiple resignations and prosecutions.

Although the Panama Papers dataset itself was not open data, published by a government, the investigations that followed demonstrated the investigative potential of corporate ownership data - and the value of having linked and structured data, as opposed to just documents.

Crucially, open data did play an important role in the follow up Panama Paper news stories. Open Corporates, who host open data on millions of companies and shareholders worldwide, reported a substantial spike in searches from countries where political leaders were implicated in offshore company scandals - revealing citizen interest in finding out more about their politicians business dealings.

Prevention

Prevention involves actions, mechanisms and tools that reduce corruption risks, or increase the costs of corruption in ways that deter corrupt activity. Open data can be used in prevention across a range of sectors.

In the private sector

In the financial sector, governments have increasingly introduced **Know Your Customer** (KYC) regulations that require banks and other financial institutions to conduct due diligence checks what taking on a new client, or processing funds. Firms entering into new deals may also wish to carry out due diligence on potential business partners. These checks often involve:

- Identifying the owners and beneficial owners of a company client;
- Checking clients and their owners against a list of politically exposed persons or public officials;
- Checking clients and their owners against court records;

At present, these checks are often carried out using 'black box' private due diligence services. These services are often expensive (banks may pay-per-search), and often rely on a limited range of sources, such as media coverage, to flag up potential client risks. If the media have not reported on a given corruption case in the past, the due diligence databases might have a blind spot.

As more information becomes available as open data, and if government and private firms demand better quality due diligence, then there is scope for innovation in how these processes take place. However, it is important to note that many regulations and businesses processes for due diligence rely on **documentary evidence**. An entry in a dataset may need to be backed up by other sources of evidence before a bank or financial institution will make a due diligence decision based on it.

Case study: Open Ownership

The OpenOwnership.org project is working to build a global database of beneficial ownership data, drawing on existing published data, and self-submitted information from companies and beneficial owners.

Beneficial ownership refers to "the natural person or persons who ultimately owns or controls a customer and/or a natural person on whose behalf a transaction is being conducted. It also includes those persons who exercise ultimate effective control over a legal person or arrangement."

Knowing the beneficial owner(s) of an asset is vital to be able to truly follow the money, and see through layers of shell companies and complex ownership structures. Some jurisdictions are now introducing registers of beneficial ownership, requiring companies and land registrations to provide details of their ultimate beneficial owners.

By combining data from different national registers, using a common Beneficial Ownership Data Standard (currently under development) and allowing self-submission of data OpenOwnership it is aiming to provide a ready-to-use source of information for due diligence.

The project is being developed as a multi-stakeholder partnership involving Transparency International, One, the Open Contracting Partnership, the World Wide Web Foundation, Global Witness and The B Team.

Sources: Guidance on Transparency and Beneficial Ownership, Financial Action Task Force, October 2014, http://www.fatf-gafi.org/media/fatf/documents/reports/Guidance-transparency-beneficial-ownership.pdf

Interview with Chris Taggart, from OpenCorporates.

OpenOwnership.org

In the public sector

When interest and asset disclosures are filed on paper, it can be easy for politicians and officials to leave out certain disclosures, and simply hope these will not be detected. But when disclosures are published as structured data, it becomes easier to cross-reference between the information provided in a disclosure, and the information held in other sources, such as the company register.

This increase the complexity and costs of hiding information, and creates a pressure for more accurate disclosures.

Case study: ProZorro

Ukraine's public procurement system was once notorious for corruption and inefficiency. Since launching ProZorro, the country's open source, open data e-procurement system the government has saved 14% on its planned spending (more than 300 million Euros) and seen a 50% increase in companies bidding for contracts - helping build business and citizen trust in the government process.

The Anti-Corruption Open Up Guide

Case study: 3 x 3

Proactive publication of structured data about interests and assets is relatively rare. More often, declarations are made on paper forms, or hosted on scattered websites of each institution. To respond to this challenge in Mexico, Transparencia Mexicana and partners launched the '3 of 3' campaign calling on politicians and public officials to publish three key declarations using structured templates and covering their:

- Statement of assets;
- Interest declarations; and
- Tax returns

At the moment, officials need to provide this information directly to the 3 of 3 campaign, who then re-publish it in semi-structured forms. However, the information is captured using Excel templates, offering opportunities for further analysis and cross-linking of declarations.

The Anti-Corruption Open Up Guide

Detection

With thousands of procurement processes taking place every month, and hundreds of spending transactions by governments every day, it is effectively impossible to audit every one of them manually for signs of corruption. But with structured open datasets, large-scale analysis can be carried out on a rolling basis

A common approach is 'Red Flag Analysis'. Here, a set of indicators are designed, that can be assessed either using a single dataset (e.g. procurement data), or a collection of joined-up-datasets (e.g. company registers, asset registers and spending data). Software is created or configured to then read through incoming data, and analyse activities against potential indicators of corruption. When a certain threshold is hit, users of the system will be notified by alerts, or through a dashboard, that there are cases in need for deeper investigation. This red flags approach does not prove corruption is taking place - but they highlight areas which may, statistically, be subject to higher corruption risk. This can help in targeting scarce investigatory and enforcement resources.

The Open Contracting Partnership has been leading work to develop a common framework of red flags, and to assess which fields from the Open Contracting Data Standard (OCDS) are required in order to be able to detect certain corruption risks.

Case study: Open Contracting Red Flags Framework

In "Red Flags for integrity: Giving the green light to open data solutions", the Open Contracting Partnership have identified a range of metrics that can be calculated from Open Contracting Data Standard (OCDS) data on public procurement processes in order to surface corruption risks. The study identifies *"a set of over 150 suspicious behavior indicators, or "red flags" [that] occur at all points along the entire chain of public procurement-from planning to tender to award to the contract, itself, to implementation-and not just during the award phase, which tends to be the main focus in many procurement processes. "*

By building on standardised open data, tools built around these metrics can be more easily applied to datasets from different countries.

Phase	Red Flag	OCDS Field(s)
Planning	Key planning documents are not provided	planning/documents/procurementPlan; planning/documents/projectPlan
Tender	Non-public bid opening	tender/procurementMethod
Tender	Failure to adequately advertise the request for bids or proposals	tender/documents/tenderNotice; tender/tenderPeriod/startDate; tender/tenderPeriod/endDate
Tender	Short notice to bidders to submit expression of interest or prepare bids	tender/tenderPeriod/startDate; tender/tenderPeriod/endDate
Tender	Failing to read out bid prices and terms at bid opening	tender/id; tender/value/amount; tender/awardCriteria; tender/awardPeriod/startDate
Tender	Vague description of the supply terms	tender/documents/technicalSpecifications
Tender	Tender value is higher or lower than average for this item category	tender/value/amount; tender/items/id
Tender	Certain line items remain in recurring contracts that have never been called for in the past, and/or which will not be called for in the future	tender/documents/technicalSpecifications ; awards/items/description
Tender	Changes to bids after other bid prices are known	tender/amendment/changes/property; tender/value/amount
Tender	Agents charge excessive fees, usually expressed as a percentage of the contract value, or overcharge for the work performed (Billing Abnormalities)	contracts/items/unit/value/amount

Example Red Flag -> OCDS Data Standard mapping

Source:
http://www.open-contracting.org/2016/11/30/red-flags-integrity-giving-green-light-open-data-solutions/

Investigation

There are a number of advantages to the use of open data as part of anti-corruption investigations, whether carried out by journalists, or by auditors and law enforcement. Open datasets are available across national borders: meaning that citizens, journalists or officials in one country can draw upon data from another easily - and without having to go through various administrative processes to access information. This may assist investigators working in risky contexts, allowing investigations to proceed without political interference, or placing a spotlight

on the investigator. It can also support easier investigation of cross-national corruption networks. For example, the coordinators of Sinar Project in Malaysia report using the UK Beneficial Ownership register to investigate the foreign company holdings of politically exposed persons from Malaysia[22].

Case study: Building investigatory tools

In Slovakia, the Fair Play Alliance has created Datanest - a platform that compiles information on government spending from a variety of sources. This information, covering government expenditure, including on subsidies, government contracts, and elections can be queried by investigative journalists, analysts, watchdog organizations, and ordinary citizens to explore potential stories, or search for evidence to back up an investigation.

Similar projects exist in Latin America, where the PODER network have built the 'QuiénEsQuién.Wiki' platform to combines procurement information and company ownership information to support journalistic investigations.

The Anti-Corruption Open Up Guide

Case study: Uncovering the Illegal Jade Trade

In October 2015, Global Witness, published a story reporting on Myanmar's multi-billion dollar jade industry. The report focussed attention on the powerful military, government and narcotics actors benefiting from Myanmar's jade wealth, and the way in which they are using a web of anonymous companies to hide their gains at the expense of the rest of the population. The report was covered in the Wall Street Journal, Guardian, BBC, New York Times, Reuters, Wired, ABC, AP, and AFP.

The investigators from Global Witness relied heavily on company data from the Directorate of Investment and Company Administration (DICA), made accessible as open data through OpenCorporates.com. This data included key fields, such as company names, directors, and unique identifying numbers for those directors. Although some source data was removed from the DICA register during the period of the Global Witness investigation, it remained accessible in the open data copy, enabling researchers to continue following up leads - combining digital analysis techniques with conventional journalistic interview practices.

Source:
https://medium.com/opencorporates/how-open-company-data-was-used-to-uncover-the-powerful-elite-benefiting-from-myanmar-s-multi-1ef35f88d6bd

The Anti-Corruption Open Up Guide

[22] Source: Personal conversation, OGP Summit 2016.

Enforcement

There are three main use-cases for open data when it comes to enforcement. Firstly, trend analysis with open data can be used to target scarce enforcement resources: highlighting emerging issues, or surfacing areas where there is a good possibility of successful prosecutions. Secondly, open data of all forms can be part of the evidence in a case around corrupt activity. Thirdly, open data on courts, enforcement and sanction processes can be used to scrutinise the effectiveness of the enforcement system itself, and to highlight areas in need of systemic improvement.

Box 7: The importance of evidence

Ahead of the UK Anti-Corruption Summit in May 2016, we held a workshop with investigative journalists and law enforcement practitioners to explore the potential uses of open data. This workshop explored how:

- Open data can help civil society and media to put corruption issues on the law enforcement agenda, particularly when law enforcement has limited time and resources;
- In general, corruption-focused law enforcement have more referrals than they can process. Whilst some agencies might look at bulk data analysis to identify crimes (e.g. cyber crime, child exploitation), this is presently not as common in fraud and anti-corruption work;
- Law enforcement are governed by very strict rules framing how evidence can be gathered and used, and requiring that there is a replicable process, carried out by experts, between any original source of evidence, and any analysis that might be presented to a court;
- Defence lawyers won't challenge the evidence – they will challenge the process of evidence collection and processing.

This raises some key challenges for open data use that ultimately aims to secure convictions or sanctions through legal process. Although the use of open data is generally exploratory, it is important in working with open datasets to track their provenance and address the 'four Cs of evidence':

- **Context.** Where has the data come from? How was it acquired? Law enforcement cannot use any data that was illegally obtained.
- **Corroboration.** Data on its own is not enough. Behind every data is a human, and it is important to prove what human actions were.
- **Currency.** How up to date is the information? For example: There are different filling times across company registries and information can become out of date quickly.
- **Completeness.** Has the data been changed in any way? If so the defence will want to be

able to follow the same trail and mirror it. How can we be sure that this is a complete picture that someone else can replicate?

Source: Open Data, Investigations and Law Enforcement Workshop, April 2016, London.

Section 4: Conclusions

Open data is a crucial tool in addressing complex corruption networks. There is promising emerging practice, and examples of impact. However, there is a long road ahead to ensure transparency and anti-corruption policies all lead to interoperable data that is useable and effective.

Conclusions

This first version of the Anti-Corruption Open Up Guide highlights the breadth of data that can be used to address corruption networks.

In some areas, such as contracting, and beneficial ownership, strong progress towards standardised, interoperable and open data is being made. In other areas, such as land ownership, or interest disclosures, there are fewer standards or global initiatives to create such datasets.

This highlights the need for governments to move beyond high-level open data commitments, to focus on implementation and publication of specific datasets. The list in Chapter 2, and the further detail in Appendix 1 should help guide this.

It also highlights a need for multi-stakeholder collaboration around standards for key datasets. This is an important area for work by supporters of the International Open Data Charter.

Charter Open Up Guides are living documents. Over the coming year we hope this document will be updated with further details and case studies, showing the impacts that emerge when activists, governments, journalists and the private sector can put their effort into data use, rather than data gathering.

Get involved in shaping the guide further at http://www.opendatacharter.net or get in touch with the Charter team: info@opendatacharter.net.

Appendix 1: Datasets in detail

For an interactive version of this appendix go to: https://airtable.com/shrHY9KFJ5bircwvx.

Lobbying registers	**Register**
Data that allows to identify and describe individuals and organizations that act as lobbyist (i.e. represent or act in behalf certain group).	**Key features** Organisation identifier, Government party identifiers, Identifiers for individuals **Related datasets:** Public officials register, Politically exposed people's list, Meeting records, Interest declarations, Company register, List of government contractors, Contracts register
Standards: Popolo	

Interest declarations	**Register**
Data that identifies those private activities or relationships of public officials and figures in public life, which could interfere with the mandate, activities or decisions of public nature made them.	**Key features** Identifiers for individuals, Company identifier, Organisation identifier **Related datasets:** Land register, Company register, Government grants, Contracts register
Standards: Popolo	

Company register	**Register**
Dataset that identifies every company legally registered to operate within a jurisdiction.	**Key features** Organisation identifier, Beneficial ownership, Company identifier **Related datasets:** Interest declarations, Licenses, List of government contractors, PPPs, Contracts register, Public procurement complaints register
Standards: OpenCorporates Schema, EITI Standard	

Charity register	**Register**
Data repository about every charity or non-for-profit legally registered to operate within a jurisdiction and that could access to specific fiscal or tax status.	**Key features** Organisation identifier **Related datasets:** Political party finance, Government grants, Meeting records, Interest declarations
Standards: *None identified*	

Corruption-sensitive posts	**Register**
List of positions within government that based on their mandate can be prone or sensitive to corruption, such as public procurement.	**Key features** Identifiers for individuals **Related datasets:** Politically exposed people's list, Interest declarations, Assets declarations, Budgets, Government grants, Contracts register, Lobbying registers
Standards: Popolo	

Politically exposed people's list	**Register**
A list of individuals who are or have been entrusted with a prominent function (FATF, 2013).	**Key features** Identifiers for individuals **Related datasets:** Corruption-sensitive posts, Interest declarations, Assets declarations, Contracts register
Standards: Popolo	

Public officials register	**Register**
List of all public officials above a certain level of seniority, along with details of their role. This might also extend to cover government advisors.	**Key features** Identifiers for individuals **Related datasets:** Corruption-sensitive posts,

Standards: Popolo	Assets declarations, Interest declarations, Meeting records

List of government contractors	**Register**
Register of all the contractors who hold or have held contracts with government, and details of any sanctions they have received.	**Key features** Identifiers for individuals, Contract identifiers, Company identifier **Related datasets:** Contracts register, Interest declarations, Licenses, PPPs, Public procurement complaints register
Standards: Popolo	

Government council or advisory boards register	**Register**
A list of all the existing advisory councils and boards to government, including data about their mandate and members.	**Key features** Identifiers for individuals, Dates, Name, Number of meetings **Related datasets:** Politically exposed people's list, Government grants, Contracts register, Voting records
Standards: Popolo	

Political party finance	**Public disclosures**
Records on financial contributions and any other type of contribution received by a politician or a political party during a period of time, as well as their spending.	**Key features** Identifiers for individuals **Related datasets:**:Related datasets
Standards: Popolo	

Budgets	Public disclosures
Data on the expected income and expenditure a government or an specific government during a period of time.	**Key features** Values, Government party identifiers **Related datasets:** Spending, Register of government projects
Standards: Fiscal Data Package, EITI Standard, IATI	

Tender and award processes	Transactions
Data on advertised opportunities for public contracts, tendering procedures and awarded contracts.	**Key features** Organisation identifier, Contract identifiers, Company identifier, Paid amounts **Related datasets:** Contracts register, List of government contractors, Company register
Standards: Open Contracting Data Standard	

Licenses	Transactions
Register on the permissions that have been issued by a government to engage in certain business or perform a specific activity.	**Key features** Organisation identifier, Contract identifiers, Company identifier **Related datasets:** Land register, Budgets, Contracts register, Company register, Tender and award processes
Standards: Open Contracting Data Standard, EITI Standard	

PPPs	Transactions
Data about every public-private partnership that have agreed by a government or that is under planning, that allows to develop or implement a public project in collaboration with a private entity.	**Key features** Organisation identifier, Contract identifiers, Company identifier, Spending records **Related datasets:**

Standards: Open Contracting Data Standard	Contracts register, Budgets, Company register, Interest declarations

Spending	**Transactions**
Records on the expenditures incurred by each government unit.	**Key features** Spending records, Government agencies initial budget, Paid amounts , Contract identifiers **Related datasets:** Budgets
Standards: Fiscal Data Package	

Government grants	**Transactions**
Refers to data about donations and other type of contributions that a government has granted to different organisations or individuals. A dataset might also contain information on decision making around these grants.	**Key features** Organisation identifier, Grants' beneficiaries, Grant's conditions and regulations, Contract identifiers **Related datasets:** Charity register, Contracts register, Interest declarations
Standards: 360 Giving, eGrant, IATI	

International aid and financing	**Transactions**
Records about resources that are given or received by a government, development agency, international organizations or civil society organization with the development and humanitarian purposes.	**Key features** Organisation identifier, Paid amounts **Related datasets:** Charity register, Assets declarations, Contracts register
Standards: IATI	

Contracts register	**Register**

A list of current and past government contracts, and details of implementation and amendments made to these over time.

Standards:
Open Contracting Data Standard, EITI Standard

Key features
Contract identifiers, Organisation identifier, Company identifier, Values, Paid amounts

Related datasets:
Company register, Interest declarations, Corruption-sensitive posts, Audit data, Budgets, Spending

Audit data

Transactions

Data on the results of auditing and evaluation procedures conducted by external or internal government entities, such as Supreme Audit Institutions.

Standards:
:Standards

Key features
Budget balance, Number and value of new financial assets, Contract identifiers

Related datasets:
Contracts register, Spending, Licenses, PPPs, Government grants, Company register

Voting records

Public disclosures

Register on individual votes made by the members of legislative and collegiate bodies.

Standards:
:Standards

Key features
Identifiers for individuals

Related datasets:
Lobbying registers, Interest declarations, Budgets, Government grants, Records of changes in regulations

Court data

Transactions

Records on legal decisions made by the judiciary, as well as on cases prosecuted.

Key features
Case records, Appeals records , Trial identifiers

Related datasets:

Standards: :Standards	Interest declarations, Politically exposed people's list, Audit data, Assets declarations, Public procurement complaints register

Register of government projects	**Register**
A register of major government projects to be implemented or that are being implemented. For example, infrastructure or educational projects.	**Key features** List of projects, Budget estimations **Related datasets:** Interest declarations
Standards: :Standards	

Meeting records	**Public disclosures**
Records and minutes of meetings held by senior public officials, specially in corruption-sensitive postings.	**Key features** Number of meetings, Meetings duration **Related datasets:** Lobbying registers, Interest declarations, Corruption-sensitive posts, Licenses, PPPs, Contracts register, Voting records, Records of changes in regulations
Standards: :Standards	

Records of changes in regulations	**Public disclosures**
Records of modifications made to legislation and other regulations, within a jurisdiction, including information on the dates they were made and their proponents.	**Key features** Dates, Identifiers for individuals, Organisation identifier **Related datasets:** Lobbying registers, Budgets, Interest declarations, Voting records
Standards: :Standards	

Campaign promises	**Register**
Register of the promises made by politicians during electoral campaigns, in order to monitor its implementation.	**Key features** :Features **Related datasets:** Contracts register, Interest declarations, Budgets, Company register, Government grants
Standards: :Standards	

Debarred or sanctioned contractors	**Register**
Register of private entities or individuals that have been debarred from participating in government contracting procedures or that have been sanctioned due to their behavior in past contracting procedures.	**Key features** Organisation identifier, Identifiers for individuals, Company identifier **Related datasets:** Interest declarations, Contracts register, Tender and award processes, Public procurement complaints register, Audit data
Standards: Open Contracting Data Standard	

Public procurement complaints register	**Register**
Register of complaints filled out by companies during public procurement procedures and information about its investigation and results.	**Key features** Contract numbers **Related datasets:** Tender and award processes, Contracts register, Interest declarations
Standards: Open Contracting Data Standard	

Land register	**Register**
Registers of land and property owners within a jurisdiction.	**Key features** Beneficial ownership, GeoData

Standards: *None identified*	**Related datasets:** Assets declarations, Licenses, Budgets, Register of government projects

Tax records	**Transactions**
Registers of taxes and other fees or payments, such as royalties, collected by government and state-owned enterprises.	**Key features** Identifiers for individuals, Paid amounts , Company identifier, Organisation identifier **Related datasets:** Assets declarations, Contracts register, Licenses
Standards: *None identified*	

Assets declarations	**Public disclosures**
Information and data about the assets held by public servants and relevant politicians and their relatives. In many countries disclosing assets is mandatory for any public servant.	**Key features** Identifiers for individuals, Assets value, GeoData, Values **Related datasets:** Tax records, Corruption-sensitive posts, Politically exposed people's list, Public officials register
Standards: *None identified*	