

The Anti-Corruption Open Data Package

Public draft open for comments

An initiative by the



open data charter

Partners



The Anti-Corruption Open Data Package

This practical resource provides references to identify the priority datasets, open standards and open data use-cases that governments, civil society and other stakeholders should focus on, in order to tackle corruption at all levels and to respond to increasingly complex **corruption networks**.

Contents

[Contents](#)

[Section 1: Analytical framework](#)

[Meeting the challenge of corruption](#)

[Linking open data to anti-corruption strategies](#)

[The Public Policy Cycle](#)

[The Anti-Corruption Cycle](#)

[Increasing responsiveness](#)

Public draft open for comments

[The role of open data in the anti-corruption policy cycle](#)

[Section 2: Anti-Corruption Open Data](#)

[Data against corruption networks](#)

[Registers](#)

[Transactional Data](#)

[Public disclosures](#)

[Availability challenges](#)

[Foundations of a solid anti-corruption data infrastructure](#)

[Section 3: Anti-Corruption Open Data Cases](#)

[Prevention cases](#)

[Detection cases](#)

[Investigation cases](#)

[Opening anti-corruption investigations \(this case is under development\)](#)

[Enforcement cases](#)

[Open Medicines, Honduras](#)

[Annex 1, Anti-corruption datasets in detail](#)

Section 1: Analytical framework

Public draft open for comments

Whilst recognizing the importance of transparency to inhibit corruption, it is key to go beyond the idea that disclosing data directly equals reducing corruption. **Open data** can play a key role to dismantle corruption networks, if governments secure its availability and interoperability.

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 systems¹. As a global reference – and as a starting point for this resource– **Transparency International** defines corruption as “the abuse of entrusted power for private gain”². 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 categories³ that allow to group and analyze certain acts of corruption:

1. **Grand corruption** consists of acts committed at a high level of government that distort policies or the central functioning of the state, enabling leaders to benefit at the expense of the public good.
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.

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 are 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 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

¹ 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.

² Transparency International. Available at: <https://www.transparency.org/what-is-corruption#define>

³ Idem.

Public draft open for comments

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 collaboration⁴. 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)⁵. 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.
- 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”⁶. 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 different actors.

⁴ G20. Available at: <http://www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf>

⁵ Transparency International. Available at: http://www.transparency.org/exporting_corruption

⁶ United Kingdom Government website. Available at: <https://www.gov.uk/government/topical-events/anti-corruption-summit-london-2016/about>

Public draft open for comments

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 to adjust and improve different anti-corruption strategies faster, as well as to 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**”⁷. 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*”⁸.

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**⁹.

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

⁷ International Open Data Charter. Available at: <http://opendatacharter.net/principles/>

⁸ Ídem

⁹ 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>.

Public draft open for comments

By acknowledging the complexity of this discussion, the **Anti-Corruption Open Data Package** 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 Data Package 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.

Box 4. Developing the Anti-Corruption Open Data Package

The Anti-Corruption Open Data Package 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

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.

[Track 3] Acknowledging anti-corruption efforts using open data: relevant experiences

[Track 2] An initial input: key anti-corruption data

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 4] Frontiers, challenges and ideas: potential uses of open data for anti-corruption

Public draft open for comments

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, where gaps in current data availability were identified, a review was carried out of shared corruption problems experienced in different countries that might be addressed with open data in future. These sources of information allowed to develop potential uses cases of open data for anti-corruption, looking ahead to the possibilities that should be unlocked when more data is published.

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 cycle of certain public policy, 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 and modalities involved in tackling corruption (*see table 2*). Both cycles are intrinsically related and interconnected since any public policy can be affected by corruption, as entrusted power can be abused for private gain in different moments of the public policy in question.

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
-------------	--------------------------	-------------------------

Public draft open for comments

		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 Data Package		

The Anti-Corruption Cycle

The **anti-corruption cycle**¹⁰ 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-	Definition	Potential actors involved
---------------------	------------	---------------------------

¹⁰ Consultation for the Anti-Corruption Sector Package. Available at: <https://goo.gl/GLtvFt> - building on the G20 Anti-Corruption Working Group open data principles.

Public draft open for comments

corruption cycle		
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 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 ¹¹ and compile information about the illicit behaviour detected and the parties involved.	CSO's, investigative journalists, oversight institutions (auditors, comptrollers, parliament), national and foreign 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 Data Package		

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.

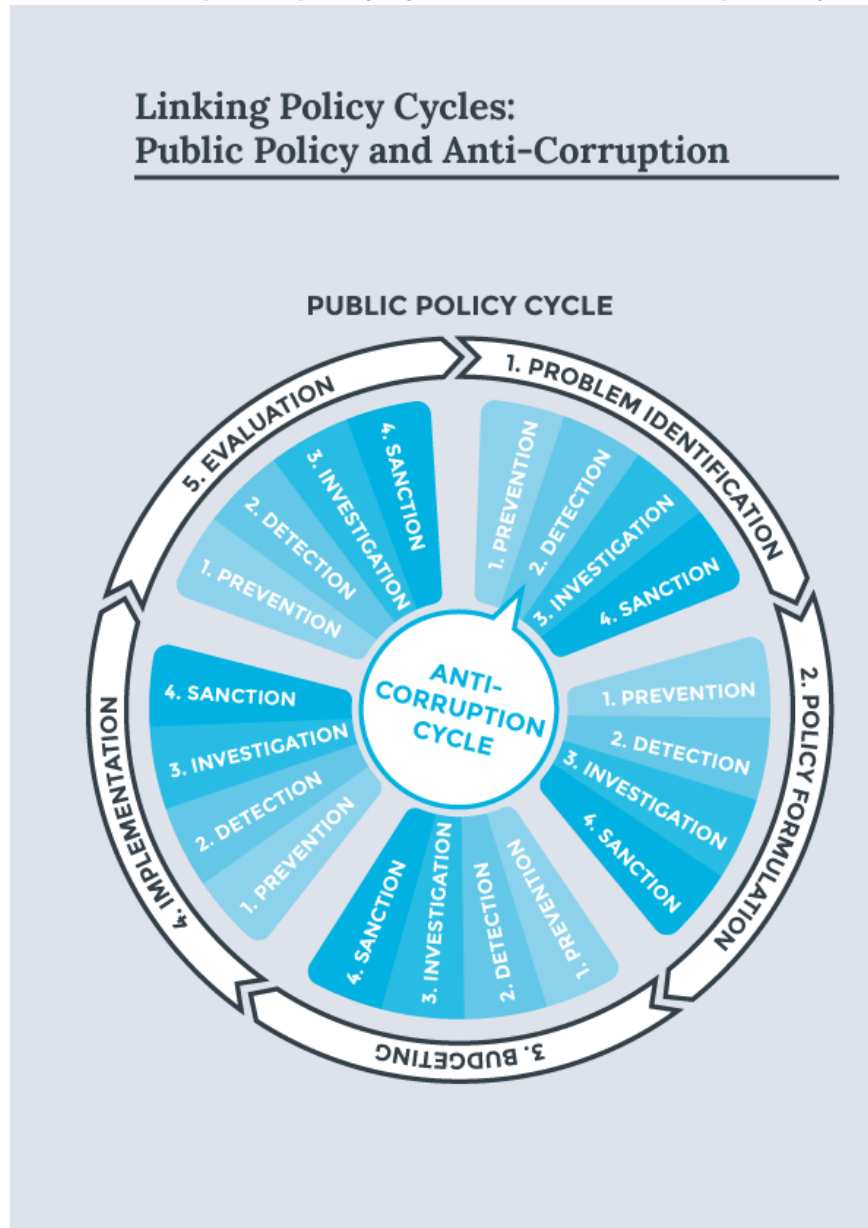
A specific public policy, considered against the goals that were set out for it, may be efficient and effective, and yet at the same time be affected by vested interests or even by corruption. For example, when analyzing a corruption case, an investigative journalist may focus on the budgetary stage 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.

¹¹ 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.

Public draft open for comments

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 Data Package** 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 anti-corruption strategies can make the implementer actor –whether it is government, civil society or journalists– more responsive to corruption risks and cases, through the use of open data.

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 tasks can happen at the same time.
For example, during the problem identification of a public policy, corruption can be detected.*

Public draft open for comments

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 articulated 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

Public draft open for comments

Enforcement	Integration of legal records and court documents.	Achieve effective sanctioning and assets recovery.	Oversight institutions (auditors, comptrollers), prosecuting institutions, judiciary.
The Anti-Corruption Open Data Package			

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 the 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 with other databases, cross-country comparisons or complexer anti-corruption investigations.

Section 2: Anti-Corruption Open Data

A solid **anti-corruption data infrastructure** can only be built when it is guaranteed that the different datasets that are part of it can communicate with each other. The adherence to **open data standards** contributes in securing that a larger number of users can benefit from the data available.

Data against corruption networks

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*).

The approach to make sure the data identified here can be available as open data 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, these datasets form a basic core that countries should strive to make available and interoperable. Together, these datasets can form a solid **anti-corruption data infrastructure**.

This data takes 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.

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.	<ul style="list-style-type: none"> – 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	<ul style="list-style-type: none"> – Budgetary datasets – Government spending – Contracts – Public-private partnerships – Political financing – Licenses and permits – Grants / scholarships – Auditing datasets

Public draft open for comments

	resources.	
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.	<ul style="list-style-type: none"> – Voting records – Meeting records – Court records – Campaign promises
Rent extractions	Refers to any dataset containing records and information on the destiny of public 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.	<ul style="list-style-type: none"> – Assets declarations/registers – Land and property registers – Tax databases
The Anti-Corruption Open Data Package		

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”¹². 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’¹³.

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 Data Package

¹² <https://gds.blog.gov.uk/2015/09/01/registers-authoritative-lists-you-can-trust/>

¹³ <https://gds.blog.gov.uk/2015/10/13/the-characteristics-of-a-register/>

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 linked with information from a register, would show illicit benefits received by a politically exposed person.

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 Data Package

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.

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

Public draft open for comments

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
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 register	Rent Extraction	Register
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 Data Package

Is there a key anti-corruption dataset missing? Let us know! Access here

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 anti-corruption activity 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 ¹⁴	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.		
Source: World Wide Web Foundation. Elaborated for the Anti-Corruption Open Data Package		

¹⁴ 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 it is guaranteed that the different datasets that are part of it can communicate with each other. The higher the number of connections, the higher the number of possibilities of articulating data for an anti-corruption intervention. 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 can provide reference of the format a specific data field must have to secure its interoperability with other datasets. The use of open data standards allow to generate unique identifiers for individuals and organizations, set specific parameters to register events or transactions, and collect and organize data with minimum quality requirements. 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, can together review the availability of such data and agree on a route map to disclose it as open data. At the same time, it will be important to review how data is structured and assess if it is worth to adjust it to match with open data standards that will allow linking it with other data.

Table 5. Summary of priority data standards for building an anti-corruption data infrastructure		
Name of data standard	Description	Sponsor
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.	Open Corporates (Private firm)
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 is under development.	Open Contracting Partnership (CSO)
Fiscal Data Package	Schema for publishing and consuming quantitative fiscal data, especially data generated during the planning and execution of budgets. It supports both data on expenditures and revenues ¹⁵ .	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, organizations,	

¹⁵ Open Knowledge Foundation

Public draft open for comments

	motions and voting, events, speeches, among others.	
Beneficial Ownership Standard		
The Anti-Corruption Open Data Package		

Is there a key anti-corruption data standard missing? Let us know! [Access here](#)

Value for supporting countries anti-corruption commitments. 1K4AC.

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

The G20 has recently pushed the open data data globally. Accounting for 85% of the gross world product (GWP), 80% of world trade and two-thirds of the world population¹⁶, actions implemented by these countries can detonate trends across the world. Taking it 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 portals¹⁷ 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.

The Anti-Corruption Open Data Package

Box 7. What is Beneficial Ownership?

¹⁶ OECD 2015 numbers

¹⁷ The G20 Open Data portals were reviewed in april 2016

Public draft open for comments

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.

Identifiers and company names are both minimum information that should be available for the public. Beneficial ownership datasets should contain legal name, legal address, date of incorporation, type of company (limited or unlimited liability), track record of directors, and any details of the nature of that ownership and control. When the highest level of ownership that can be found is a company, the data might include this corporate parent information with an organisation identifier for the company.

Source: *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.

Section 3: Anti-Corruption Open Data Cases

INTRODUCTION

Prevention cases

[Under development]

Detection cases

[Under development]

Investigation cases

Opening anti-corruption investigations *(this case is under development)*

Investigation might be the most complex stage of the anticorruption cycle, as it implies gathering enough information and evidence to support the initial claim and start a formal sanctioning procedure, such as a trial. Specifically, a corruption investigation should be able at least to:

1. Identify each individual member involved in the corruption network
2. The degree of involvement of each member
3. The type and/or amount of benefit obtained by each member
4. The techniques and agreements used to commit a corruption crime

In other words, an effective investigation will be able to describe the full “modus operandi” of a corruption network. An investigation can be conducted either by social actors –such as journalists and civil society organizations– or by government investigation agencies and prosecutors. The main difference among these two types of investigations is that only government investigations can lead to legal sanctions. Even when social investigations might be useful to increase public awareness, they will mainly lead to social sanctions. Only under certain legal circumstances, social investigations can lead to formal government investigations. For example, the Foreign Corrupt Practices Act (FCPA) in the United States allows both, the Department of Justice (DOJ) and the Securities and Exchange Commission (SEC), to start a formal investigation when a corruption crime is exposed in the media.

Taking into consideration the link between corruption investigations and impunity –which was highlighted in one of the workshops to develop this resource–, this case will only focus on government investigations. Despite investigation agencies and prosecutors have access to a wide range of investigation tools –including access to either public and private information systems and databases– the generation of solid evidence is still seen as a challenge for effective sanctioning.

Public draft open for comments

Acknowledging the need of **secrecy** that a government investigation might need to avoid affecting the results of the investigation, destruction of evidence and due process, this case will address the use of data from a different perspective. Instead of identifying which data can be used to generate evidence about specific corruption crimes, this case will focus on how open data can increase (1) openness of corruption investigations and (2) accountability of prosecuting agencies.

Enforcement cases

Open Medicines, Honduras¹⁸

Honduras faces the highest level of economic inequality in Latin America, with more than 63 percent of the population living in poverty in 2014. Since the 2008 global economic crisis, the country has experienced a moderate recovery, with GDP rising to 20.15 billion USD in 2015. However, despite this growth, the country still faces resource constraints that take a toll on health spending.

This is especially concerning as most of the Honduran population depends on state-delivered public health services, including the provision of medicines and other treatments by the government at no direct cost. However corruption –the country’s Corruption Perception Index Ranking is 112 out of 168 countries, with an overall score of 31 out of 100– resulted in powerful and well-connected pharmaceutical companies inflating prices and forcing the purchasing of non-essential or low-quality medicines. With the limited health budget of the Government of Honduras being used so ineffectively, the result was less essential medicine available to the public.

While this issue was generally understood by civil society and development partners, a 2010 dengue fever epidemic across Central America highlighted the problem, as Honduras endured far greater numbers of casualties than neighboring countries because of a lack of saline solution.

The goal of **Open Medicines** is to promote greater efficiency in the use of public funds used to purchase medicines for the country’s public health system, addressing the chronic shortages currently faced in Honduras. The case falls under the **enforcement** phase of the Anti-Corruption Cycle, which refers to mechanisms intended to prosecute and penalize corruption crimes, as well as to recover illegal rents stolen or generated by a corruption crime.

The Asociación Para Una Sociedad Más Justa (ASJ) –Transparency International’s chapter in Honduras– carries out research and advocacy related to all the steps in the medicine procurement process, including registering new treatments with the authorities, ensuring delivery to the Ministry of Health, and distribution to clinics across the country. As part of this effort, Open Medicines, launched in 2014, focuses on enabling civil society to track **prices**, **suppliers**, and **quantities** of medicines purchased over time and identify anomalies that may indicate corruption. The focus is on this contracting phase, as this is when public funds are spent.

Data collected relates to procurements at the Ministry of Health, as well as purchases made by the National Social Security Institute. Combined these make up roughly 90 percent of government

¹⁸ This case was developed by Piper Hart as a researcher for the Anti-Corruption Open Data Package.

Public draft open for comments

spending on medicines. Datasets include suppliers and prices and quantities of medicines purchased over time. This information is obtained in large part through the Honduran Government's transparency portal. When information is not available through the transparency portal, ASJ staff liaise directly with government officials or make requests through the Freedom of Information Act.

Extensive work is done to standardize and correct the information collected before being added to the Open Medicines platform in order to facilitate comparison and analysis. Prices are shown adjusted for inflation in order to give a real sense of changes over time. The Open Medicines website is updated roughly once per year, in line with the Government's purchasing schedule. Informal feedback from users, including the Ministry of Health and pharmaceutical companies, has been received. In the future, the platform will be updated to include communications channels between users and operators and to continue improving ease of use.

Once added to the Open Medicines website, users are able to **visualize trends across time** by selecting specific datasets and running analysis through the platform. This information is then used by law enforcement to investigate and prosecute corruption offenses.

ASJ's data analysis and advocacy efforts, including the Open Medicines platform, has helped bring to light specific cases of corruption and abuse of the medicine procurement system. For example, data assembled and visualized through Open Medicines highlighted how one pharmaceutical company went from having virtually no government contracts to becoming one of the largest suppliers to the government after one of its founders became the Vice President of Congress. The owners of this company are now facing several related corruption charges pressed by Honduras' Public Prosecutor.

Annex 1: Anti-corruption datasets in detail

Lobbying registers	Register
Details of lobbyists, who they are lobbying, who they are lobbying on behalf of and what they are lobbying about.	Key features Organisation identifier number, Government party identifiers Related datasets: Public officials directories, Politically exposed people's list
Standards: Popolo	

Interest declarations	Register
Allows the identification of those activities or relationships which could interfere with the mandate, activities or decision making of public officials or other figures in public life.	Key features Identifiers for individuals Related datasets: Land register, Company register
Standards: Popolo	

Company register	Register
List of every company legally registered to operate within a jurisdiction with company identifier and legal name as minimum information, other information should include type of company, date of incorporation, company directors.	Key features Organisation identifier number, Beneficial ownership, Company identifier Related datasets: :Related datasets

Public draft open for comments

Standards: OpenCorporates Schema	
--	--

Charity register	Register
List of 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 number Related datasets: :Related datasets
Standards: :Standards	

Corruption-sensitive postings	Register
List of names of all the public servants who are in positions where they take decisions that are prone or sensitive to corruption, such as public procurement.	Key features :Features Related datasets: Politically exposed people's list
Standards: Popolo	

Politically exposed people's list	Register
A list of individuals who are or have been entrusted with a prominent function (FATF).	Key features Identifiers for individuals Related datasets: Corruption-sensitive postings
Standards: Popolo	

Public officials directories	Register
Containing a list of all public officials above a certain level of	Key features

Public draft open for comments

<p>seniority, along with details of their role. This might also extend to cover government advisors.</p>	<p>Identifiers for individuals</p> <p>Related datasets: Corruption-sensitive postings</p>
<p>Standards: Popolo</p>	

List of government contractors	Register
<p>Register of all the contractors who hold contracts with government, and details of any sanctions they have received.</p>	<p>Key features Identifiers for individuals</p> <p>Related datasets: Contracts register</p>
<p>Standards: Popolo</p>	

Council / advisory board members	Register
<p>A list of all the members of advisory councils and boards to government.</p>	<p>Key features Identifiers for individuals</p> <p>Related datasets: Politically exposed people's list</p>
<p>Standards: Popolo</p>	

Political party finance	Public disclosures
<p>Records on financial contributions received by a politician or a political party during a period of time, as well as their spending.</p>	<p>Key features Identifiers for individuals</p> <p>Related datasets: :Related datasets</p>
<p>Standards: Popolo</p>	

Public draft open for comments

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	

Tender and award processes	Transactions
Data on advertised opportunities for public contracts, tendering procedures and awarded contracts.	Key features Organisation identifier number, Contract identifiers Related datasets: Contracts register
Standards: Open Contracting Data Standard	

Licenses	Transactions
Register on the licenses that have been issued by a government.	Key features Organisation identifier number, Contract identifiers Related datasets: Land register
Standards: Open Contracting Data Standard	

PPPs	Transactions
Information on the planning, tendering, initiation or operation of Public Private Partnerships.	Key features Organisation identifier number, Contract identifiers Related datasets:

Public draft open for comments

Standards: Open Contracting Data Standard	Contracts register
---	--------------------

Spending	Transactions
Records on the expenditures incurred by government agencies	Key features Spending records, Government agencies initial budget
Standards: Fiscal Data Package	Related datasets: Budgets

Government grants	Transactions
Containing details on the grants that government has issued to different organisations or individuals. A dataset might also contain information on decision making around these grants.	Key features Organisation identifier number, Grants' beneficiaries, Grant's conditions and regulations
Standards: 360 Giving, eGrant	Related datasets: Charity register

Contracts register	Register
A list of current and past government contracts, and details of amendments made to these over time.	Key features Contract identifiers, Organisation identifier number
Standards: Open Contracting Data Standard	Related datasets: Company register

Audit data	Transactions
-------------------	---------------------

Public draft open for comments

Data on the results of auditing and evaluation procedures conducted by external or internal government entities, such as Supreme Audit Institutions.	Key features Budget balance, Number and value of new financial assets
Standards: :Standards	Related datasets: :Related datasets

Voting records	Public disclosures
Registers on individual votes in Congress (Session, chamber, law category (amendment, new bill, nomination, etc.)). Members of congress, bills and resolutions, committees. Voting records of any other collegial body.	Key features Identifiers for individuals, Scores
Standards: Open data to legislative transparency	Related datasets: :Related datasets

Court data	Transactions
Records on decisions taken by the judiciary, such as case records, judicial records, appeals and rule decisions.	Key features Case records, Appeals records , Trial identifiers
Standards: :Standards	Related datasets: :Related datasets

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
Standards: :Standards	Related datasets: :Related datasets

Public draft open for comments

--	--

Meeting records	Public disclosures
Records and official minutes of meetings held by public officials and government bodies.	Key features Number of meetings, Meetings duration
Standards: :Standards	Related datasets: :Related datasets

Records of changes in regulations	Public disclosures
Records of changes in legislation overtime, including information on the amendments and their proponents.	Key features Number of changes
Standards: :Standards	Related datasets: :Related datasets

Campaign promises	Register
Register of the promises made by politicians during the campaign process, in order to monitor its implementation. Campaign promises should be able to link with government contracts datasets.	Key features List of promises, List of fulfilled projects
Standards: :Standards	Related datasets: :Related datasets

Debarred or sanctioned contractors	Register
Register of government contractors that have been debarred or sanctioned. Registers may include information about the cause, date and sanction imposed.	Key features Organisation identifier number, Identifiers for individuals

Public draft open for comments

	Related datasets: :Related datasets
Standards: :Standards	

Public procurement complaints registers	Register
Register of complaints filled out by companies during public procurement procedures and information about its investigation and results.	Key features List of complaints, Contract numbers Related datasets: Tender and award processes
Standards: :Standards	

Land register	Register
Registers of land and property owners within a jurisdiction.	Key features Beneficial ownership, GeoData Related datasets: :Related datasets
Standards: :Standards	

Tax records	Transactions
Registers of certain financial transactions and taxes paid by PEPs, state-owned enterprises, or companies.	Key features Identifiers for individuals, Paid amounts Related datasets: :Related datasets
Standards: :Standards	

Asset declarations	Public disclosures
---------------------------	---------------------------

Public draft open for comments

<p>Information and data about the assets hold by public servants and relevant politicians and their relatives. In many countries disclosing assets is mandatory in order to be a public servant.</p>	<p>Key features Identifiers for individuals, Assets value</p> <p>Related datasets: :Related datasets</p>
<p>Standards: :Standards</p>	