

Putting the Open Data Charter into practice

Consultation on anti-corruption resource package

This document summarises an initial plan for the development of a companion resource to the International Open Data Charter, and invites comments, feedback and contributions to shape the draft resource, to be developed over the first half of 2016.

You will find a number of consultation questions (Q) in this document. You can use the comments feature to share your responses by highlighting the question and clicking



You can also send feedback to anticorruption@opendatacharter.net.

It is based on initial work by the Government of Mexico, and a workshop held alongside the Open Government Partnership Global Summit in Mexico City in October 2015.

CONTENT

[CONTENT](#)

[INTRODUCTION](#)

[BUILDING AN ANTI-CORRUPTION DATA INFRASTRUCTURE](#)

[IDENTIFYING THE INFRASTRUCTURE: USERS FIRST](#)

[Modalities](#)

[Cross cutting themes](#)

[Datasets and data elements](#)

[Resource structure](#)

[HOW TO ENGAGE](#)

INTRODUCTION

The international [Open Data Charter](#) was launched in September 2015, and adopted by the first 17 governments in October 2015, with the objective of fostering greater coherence and collaboration for the increased adoption and implementation of open data principles, standards and good practices across sectors around the world. To support implementation of the [fourth principle of the Charter](#), on comparable and interoperable data, a series of resources are to be created, outlining key data elements and datasets states should look to provide.

The Charter has been used as a core reference work in the development of the '*G20 Anti Corruption Open Data Principles*', and thus the first of these resources will focus on open data for anti-corruption

BUILDING AN ANTI-CORRUPTION DATA INFRASTRUCTURE

Corruption can be defined as “the abuse of entrusted power for private gain”¹. Through corruption, billions of dollars of public funds are lost every year, and levels of trust in governments, societies and economies is undermined. Corruption is a problem facing both developed and developing countries, creating a drag on development and progress.

The G20 has recognised the vital role data can play in challenging corruption. As [ONE document](#), there are many cases where data has enabled citizens, watchdogs and governments to ‘follow the money’ and hold corrupt actors to account. However, the release of isolated datasets offers a hit-and-miss approach to challenging corruption. To truly enable data to act as a resource for anti-corruption action we need a joined-up data infrastructure.

Just like physical infrastructure of roads, rail and electricity supply lines, an open data infrastructure needs to follow common standards and to reach a certain level of quality before it operates effectively. In the case of data, this involves identifying the key datasets that are needed, and the key data elements that can link them up. A robust anti-corruption open data infrastructure should be able to support several kind of uses such as:

¹ Transparency International: <https://www.transparency.org/what-is-corruption#define>

- Red flag² analysis for corruption prevention and detection;
- Creation of greater transparency in procurement processes, removing opportunities for corruption;
- Easier investigations by journalists, increasing the media coverage of corruption cases;
- Streamlined enforcement investigations - reducing the reliance of officials on slow and bureaucratic interagency processes;
- International cooperation on cross-border corruption cases.

This anti-corruption data package aims to prevent, detect and deter corruption. We anticipate a positive feedback effect where the more that the data are used by key stakeholders, the more likely it is that data quality will increase: benefiting both data publishers and users.

The following section considers how such an infrastructure should be identified, building on clear use cases and user stories. The accompanying consultation invites your inputs to help move this process forward.

1K for Anti-Corruption?

Eduardo Bohorquez, director of Transparency Mexico, highlights how in looking to implement the Open Contracting Data Standard, it was important to step back and identify the 150 or so data elements that are required to have a clear view of public contracting: from dates and amounts, to descriptions of line items and procurement processes.

Eduardo has posed the question: What are the 1000 data elements needed to fight corruption? Can we identify 1k for anti-corruption? What core datasets would supply these data elements?

IDENTIFYING THE INFRASTRUCTURE: USERS FIRST

Anti-corruption action across the world draws upon hundreds of different datasets, and thousands of different data elements. However, to enable states and their partners to build a shared open data infrastructure for anti-corruption we need to prioritise. The resources accompanying the International Open Data Charter will need to present a

² Red flag approaches are usually used in public procurement and other areas as a way to detect and prevent potential corrupt practices. For more information you can find this Guide from Transparency International http://gateway.transparency.org/guides/approach/public_procurement

clear framework of data which countries can focus on providing, to promote global interoperability within the sector, and between other linked sectors, whilst also allowing local priorities and needs to be met.

It is also important for guidance to be built around clear user needs: connecting global agendas against corruption with a concrete understanding of how key actors can use open data to further their anti-corruption work.

Building on the G20 anti-corruption agenda, we identify four modalities in which open data can be used, and four key areas of sectoral focus.

Modalities

The guidance will focus on the four modalities of corruption prevention, detection, investigation and anti-corruption enforcement. Each modality may make different demands in terms of key data elements, and key approaches to providing usable open data. Keeping these different modalities in mind during the development of the resource will be vital.

| Modality | Example data users and uses | Example key data elements & features ³ |
|---------------|---|---|
| Prevention | <i>Banks and other institutions can draw upon open corporate registries, beneficial ownership registers and politically exposed persons data in order to better implement Know Your Customer and Anti-Money Laundering regulation. With open registers, civil society can check the quality of the lists being used, building trusts in the compliance process - rather than the datasets used being 'black boxes'.</i> | <i>Unique identifiers for legal entities, and clear data-driven approach to identify beneficial owners.</i> |
| Detection | <i>A procurement agency may use a list of debarred firms, bidders, prior contract performance information, and information on shareholders and politically exposed persons for red flag analysis - detecting potential cases of corrupt dealing.</i> | <i>Timely and forward looking data on contracting processes.</i> |
| Investigation | <i>Investigative journalists can draw upon open data to more quickly develop leads and</i> | <i>Comprehensive and reliable public registers</i> |

³ This is an indicative list only, and is not intended to be exhaustive.

| | | |
|-------------|--|--|
| | <i>stories, leading to an increase in the coverage of corruption cases. A common pattern may be for investigations to uncover fraudulent documents by comparing them to official records of licenses, permits or certificates.</i> | <i>of companies, licenses, contracts, permits or certificates.</i> |
| Enforcement | <i>Law enforcement from one country can use open data on companies and beneficial ownership from other countries to collect the specific facts they need to build a criminal case against corrupt persons or companies.</i> | <i>Linkage between data elements and key documents, or ability of the data to function as evidence to allow prosecution and eventually in court.</i> |

Q

- Can you suggest key **users** and key **user stories** relating to each of these modalities?

For **users** we are looking to identify groups (e.g. 'civil society corruption monitors') who could use data for prevention, detection, investigation or enforcement.

For **user stories** we are looking for short narratives (1 - 2 lines) explaining how they would use data.

Cross cutting themes

Four key cross-cutting areas for focus are identified from the G20 agenda :

- **Extractives industry transparency**
- **Construction sector transparency**
- **Public procurement and open contracting**
- **Integrity of public officials**

Two of these are high-risk sectors (extractives, construction), and two relate to public sector transparency and integrity (public procurement, integrity of public officials). Two areas that feature in the G20 agenda (beneficial ownership and public budgets) feature

in our model as key datasets, rather than cross-cutting themes⁴. We incorporate bribery into our cross-cutting theme on integrity of public officials.

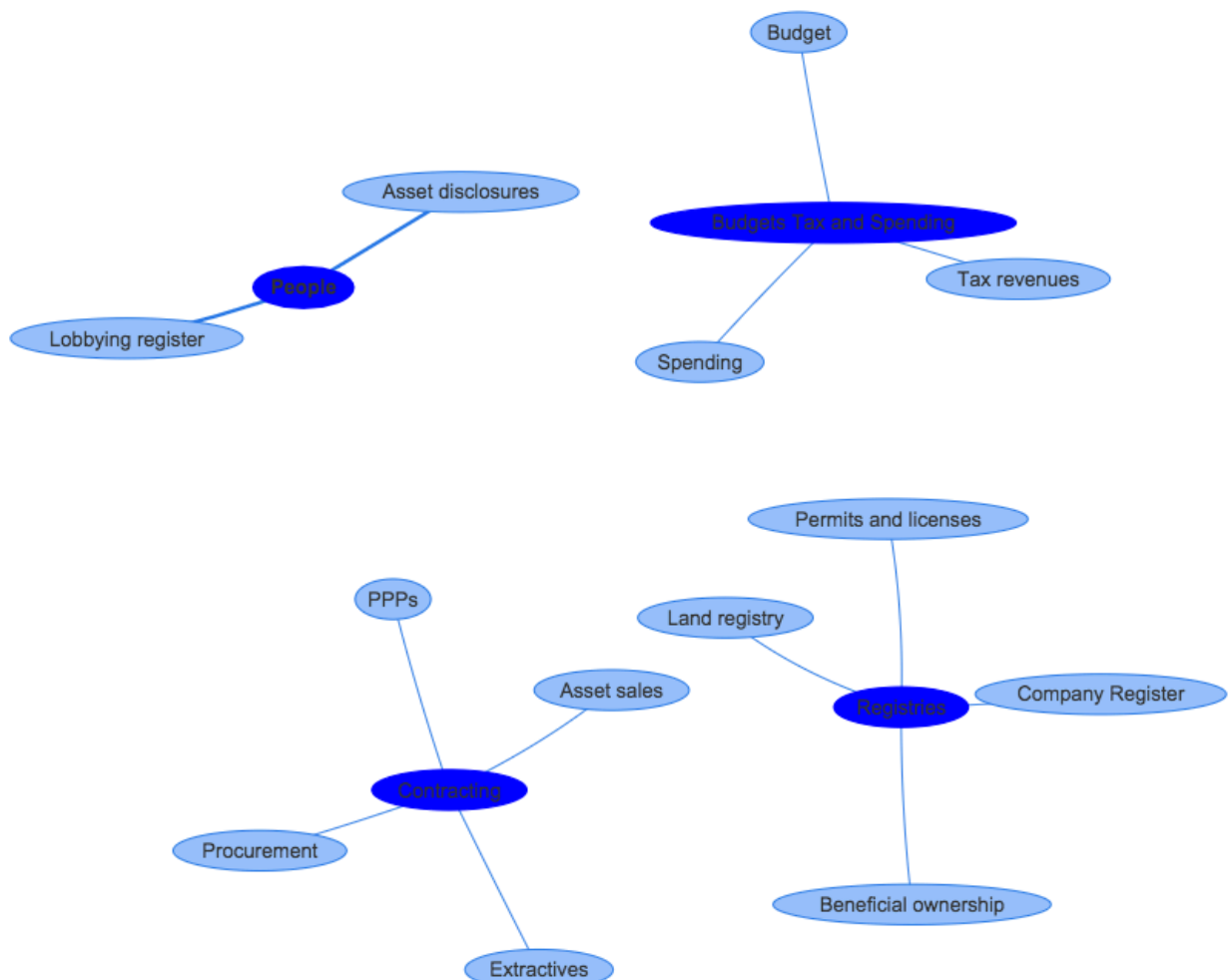
Q

- Are the proposed cross-cutting themes appropriate as a starting point for the guidance?

It will be possible to extend the guidance in future with resources for additional sectors, so the question is whether this focal set provides a feasible and relevant starting point.

Datasets and data elements

To develop this guidance will involve identifying the key data elements and datasets required as part of an anti-corruption infrastructure. The diagram below shows some of



the possible categories, datasets and data elements the guidance may focus on

⁴ An initial analysis suggests these datasets are important across each of the cross-cutting themes.

Q

- Which **datasets** and **data elements** are important for anti-corruption in each of the following areas:
 - **People**
 - **Budgets, Tax and Spending**
 - **Public registries**
 - **Contracting**

Focus on the **priority datasets** as the resource will aim to present a focussed shortlist, rather than list of all possible datasets.

Resource structure

In order to aid implementation, the resource will provide detailed pages on each of the **core data categories**. This will be made up of a number of sub-sections.

A brief *indicative example* of the kind of content this would include for a contracting page is given. In the full resource, these sections would be much more fully developed - with more supporting text and links.

| Section | Example: Contracting |
|---|--|
| <p>Problem statement & use cases: setting out the different reasons this data is important & linking to specific use-cases.</p> <p>We will draw upon existing user stories to generate the use cases, linking out to resources such as the ONE Follow the Money stories.</p> | <p>"A substantial percentage of government spending takes place through contracting, and contracts also play a key role in distributing revenues from extractives projects, or from the sale of state assets."</p> <p>Use cases:</p> <ul style="list-style-type: none"> • Procurement monitoring; • Red-flag analysis; • Modelling of extractives revenues; |

Readiness questions: highlighting key policy and technical considerations that will need to be taken into account when preparing to publish open data from this category.

The readiness questions will draw connections to wider policy considerations.

These readiness questions will not focus on the capacity to use open data: although general readiness assessment tools will be signposted which could be used to assess whether key user groups have the capacity they need to benefit from open data publication.

Key data elements & features: setting out specific individual data elements that are central to anti-corruption uses.

There may be many data elements. They will be organised in a hierarchical structure, to allow easy navigation.

Early experience implementing OCDS has shown it is useful for governments to first think through whether they collect and manage these data elements, before moving to think about specific data systems and datasets.

Key datasets: outlining the specific datasets that should be released.

The key data items that each dataset should contain will be listed also.

Procurement:

- Does the legal framework support disclosure of contracts?
- Is there an e-procurement system in place?
- Is there a threshold for information to be captured in the e-procurement system?

Extractives:

- Are extractives contracts currently disclosed?
- Is regulation on mandatory project level disclosure in place and is the data published?

- Organisation Identifiers [*]
- Planning
 - Budget Line
 - Budget
- Tender
 -

* = key joined up data item.

- **Budgets & procurement plans**
- **Tenders, awards & contracts**
- **Public spending & receipts**
- **Concession licenses**
- **Corporate ids and ownership**
- **Extractives contracts**

Standards: discussing the standards that can be used for publishing this data - and assessing the relative maturity of these standards, and examples of where they are being used.

Open Contracting Data Standard - version 1.0; open standard; developed through user-centred design process. In use in MX, CA, UA, UK.

Open Budget Data Package - alpha version, under development. Piloted in

Case studies & implementation: linking out to tools and case studies of how different countries have implemented open data in this domain. This section will focus on drawing out key lessons learned for future applications, and promote reuse and technological transfer between countries

E.g. Mexico Federal adoption of OCDS; Brazil pilot of Budget Data Package; Buy Sell Canada's OCDS implementation; Ukraine's ProZorro OCDS system; TI Slovakia's and Fair Play Alliance's mirror site of public procurement contracts.



- What content should the resource contain in each of the above sections?

HOW TO ENGAGE

Comments to be collected according to previous section structure through comments and suggestions made in this google document or by sending an email to packages@opendatacharter.net.

Deadline: 15th January 2016