

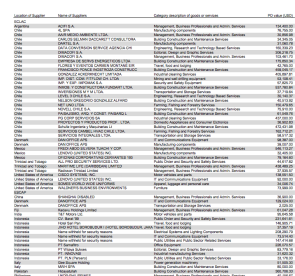
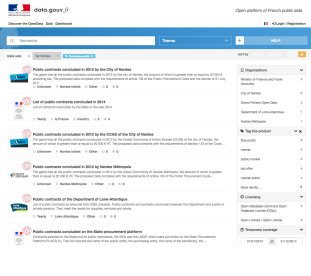


Data Guide

What data can you use for this Challenge?

Many public bodies publish *some* contracting information and data. However, not all of it is easy to access, or available as machine-readable data.

Sometimes information is only available in print, PDF, or web pages. Other times, government data portals contain exported spreadsheets and CSV files with contract data. In the best cases, you may find data structured in the Open Contracting Data Standard in JSON or CSV format.

Format	Print, PDF or websites	Structured data	OCDS format
Example	 The UN Office for Project Services publish a 300 page PDF of purchase orders and contracts.	 Data.gouv.fr has yearly exports of procurement data from national and municipal government.	 ProZorro in Ukraine provide an API and bulk download of data using the Open Contracting Data Standard JSON format.
Getting started	Convert the data into a machine-readable format using scrapers or other approaches.	Look for clear definitions of each field, and identify how the data should be interpreted.	Check the data validates against the OCDS schema. Review the OCDS documentation to understand the meaning of each field.

Note: If you find JSON structured OCDS data, but prefer working with tabular data, the [OCDS validator](#) has an option to convert valid files into spreadsheet format (and back again)

We've [put together a spreadsheet](#) listing known national sources of contracting information to help you in your search. This list is derived from the [Open Data Index](#) and [Open Data Barometer](#), which have independently collected information about procurement information available around the world. They also provide a quality assessment of this information based on a range of factors. We've calculated a bucket composite score based on the 2 subscores given to these sources by the Open Data Index and Open Data Barometer in 2015. While this bucket score should not be considered an evaluation or ranking of the quality of the data, it could act as a clue to sources with good quality data. All participants are welcome to use this list as a launching point for their data work. However, applicants should note that the ranking of these datasets is based on scoring systems developed by Open Knowledge Foundation and World Wide Web Foundation and are not reflective of OCP's assessment of their quality.

In addition, we've identified a number that are publishing data that uses the Open Contracting Data Standard (OCDS), which are listed below.

Table 1: Known OCDS Publishers

Country	Publisher	Data URL	Format	Language
Canada	Public Works and Government Services Canada	URL	JSON	English
Canada	City of Montreal	URL	CSV, JSON, PDF, XLSX	French
Colombia	Colombia Compra Eficiente	URL	CSV	Spanish
Mexico	City of Mexico	URL	JSON, HTML	Spanish
Mexico	Red Compartida	URL	XLSX	Spanish
Mexico	Grupo Aeroportuario de la Ciudad de México	URL	PDF, XLSX, JSON	Spanish
Moldova	Government of Moldova's Public Procurement Agency	URL	HTML, CSV	English, Romanian
Nigeria	Budeshi	URL	CSV, HTML	English
Paraguay	DNCP	URL	JSON, HTML	Spanish

Taiwan	DSP and National Development Council	URL	JSON	Chinese
Ukraine	ProZorro	URL	JSON, HTML	Ukrainian
United Kingdom	Crown Commercial Services	URL	JSON, CSV	English

The lists should not be considered exhaustive, and Challenge participants are welcome to use other sources, within the principles and constraints of the challenge, e.g. using at least one relevant public procurement open data source. The competitors are welcome to rely on other sources and, where suitable, produce original open data, too.

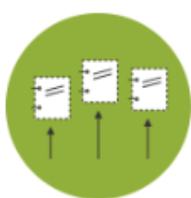
While the only requirement of the challenge is to use contracting data, participants are encouraged to link these contracting data with external sources, such as budget data. You may find it useful to look at other data sets reviewed in the Open Data Index and Barometer, as well as the [datasets](#) collected by the Open Data Charter as part of their [Open Up Guide to Anti-Corruption](#).

What kind of information is in these dataset? Will it meet your data needs?

The contracting process should include information related to multiple stages:



Planning



Initiation (Tender)



Award



Contract



Implementation

	Organisation
	Amount
	Item
	Period
	Document
	Milestones

During the planning stage, we should expect some information about the identified need. During tendering, we would expect to know what is being purchased, the timelines, the specifications. At Award, we would expect to know who won and who lost, the reasons why, and the price. The contract itself should set out the final terms of the work to be completed, the price details, the duration. During implementation, the contract may be extended, the value could go up, or the scope of work could change.

While the Open Contracting Data Standard presents an approach to publishing structured, standardized information about the entire contracting process, the current situation is that most datasets cover only one or two stages of the contracting process. For example, the UN Office for Project Services example above only has information on awarded contracts (and purchase orders), but no information about the tendering process. You could use this kind of data to focus on summary statistics, and information on the organisations awarded contracts.

Other datasets will contain enough information to allow you to focus on the process, linking tender, award, contracts and implementation. Sometimes this information is controlled by different ministries and housed in different systems. Sometimes you can ask for this information via Access to Information requests.

Finally, sometimes key fields will be missing. Check carefully to see what the datasets you plan to use contain, and how complete the fields are in your dataset (e.g. are supplier names and identifiers *always* filled in, or only sometimes?)

As can be expected from systems as complex and challenging as public procurement, the quality of data available from country to country can vary drastically. Some factors to consider when assessing the quality of these procurement datasets are:

- Phases: How many 5 phases of the procurement process (planning, tender, award, contract, implementation) are represented in the dataset?
- Agency coverage: How many agencies are included in the dataset? Are only select departments or agencies' procurement procedures included, or does the dataset contain information from every single agency?
- Tender type coverage: Which tender types are included in the datasets? Are only competitive tenders represented? Are both above and below threshold tenders included?
- Time: How many months or years of data are included in the dataset? How timely is the data (do we have close to real time information or are we looking historically?)
- Quality: Are the data properly formatted? Are name, date, and time formats standardized? Does each tender have a unique identifier? Are there obvious errors in the data entry?

In practice, there is no "perfect" dataset. All datasets will have some complications, from data gaps to poor timeliness to coding errors. That does not, however, mean that these datasets are not extremely valuable; real data sleuths can work around issues of data quality and completeness and make beautiful products out of flawed datasets. We challenge you to not only focus on the few shiniest examples, but also to consider how to work with some of the more complicated datasets, which are more representative of what one would see globally.

Tip: When data are available in multiple formats, sometimes they don't have exactly the same fields of information in each. If you are missing key information in the data files with which you are working, check web portals and other places to see if it might be available through another source.

I: The Value of Standardized Open Contracting Data

The more standardized the open contracting information, the easier it is to compare projects. We can run all sorts of analyses with a big enough sample size of comparable procurement data. Publishing and using structured and standardized information about public contracting can help stakeholders to:

- Deliver better value for money for governments: Open contracting data can help achieve good value for money (VfM) on goods and services in the procurement process, and can also help identify whether VfM has been achieved in concluded contracts. Data users want to analyze trends in prices and supplier performance, including in terms of quality and duration. Comparable data using common codelists and the availability of unit prices are particularly important for VfM use cases.
- Create fairer competition and a level playing field for business, especially smaller firms: Open contracting data can be used to understand the potential pipeline of procurement opportunities. It is a core principle of open contracting that information should be made available at the early stages of a contracting process, including information on planned procurement and invitations for tenders. Information on past contracts can allow firms to identify upcoming opportunities for re-contracting and support a more competitive marketplace, as transparency creates a level playing field with information on pricing, contract dates and key deliverables. Forward looking and timely information is particularly important for data users, as well as the ability to uniquely identify procuring entities, geographic locations, sectors, and the kinds of items which are being procured. For example, users can analyse how competitive the market is by looking at the total share of contracted dollars allocated to small firms.
- Drive higher-quality goods, works, and services for citizens: Data users want to ensure that public contracting delivers value to citizens in terms of quality of goods, works, and services provided. Monitoring contracting effectively involves being able to link budgets and development partner data to contracts and results. It also involves being able to verify whether results are being delivered on the ground. Data concerning budgets, delivery location, and subcontracting arrangements are particularly useful for those involved in contract monitoring. Using these, we can sort through implementation data to find out if what is delivered is of good quality.
- Prevent fraud and corruption: Users can run red flag analyses to find indications of suspicious or abnormal behavior. This can enable stakeholders to identify and combat corruption in public contracting. Open contracting data can be used to scrutinize procurement documents and data for signs that might indicate public

monies are being misused. There are two main approaches to fraud and corruption monitoring: a 'micro' approach that closely scrutinizes individual procurements; and a 'systemic' approach looks for suspicious patterns, and makes links between datasets to map out networks of funding, ownership and interests. Data that can be linked using globally unique identifiers for companies is particularly important for fraud and corruption detection.

- Promote smarter analysis and better solutions for public problems: Users can analyze the efficacy and efficiency of projects to find opportunities for cost savings. Public access to open contracting data builds trust and ensures that the trillions of dollars spent by governments result in better services, goods, and infrastructure projects.

The Open Contracting Data Standard (OCDS)

The power of standardized data in public procurement is immense: when data and documents on contracting are available in a structured, re-usable form, new opportunities for analysis and engagement are unlocked. Standardized data allow us to not only compare projects and unique procurement ecosystems, but also run more robust and more valid analyses of patterns within the same procurement ecosystems over time. Given the power of standardized data, the Open Contracting Partnership has developed the OCDS. The OCDS is a powerful tool to track contracting processes, and to gain insights into what is going on inside them.

We created the OCDS as a [global, non-proprietary data standard structured to reflect the complete contracting cycle](#). The standard enables users and partners around the world to publish shareable, reusable, machine readable data, to join that data with their own information, and to create tools to analyze or share that data.

Open contracting, and the data standard, covers the full contracting process, including planning, tender, award, contract and implementation. The data standard was designed and developed through an open process. It is focused on connecting up the data or documents that governments collect with the needs of users who want to help fix problems, analyze public contracting, and innovate the way contracts are made and delivered.

Though participants in this challenge are not required to use the OCDS in their entries, doing so is highly encouraged. We believe that using the standard to help organize and use your data will help make your idea more scaleable and adaptable to other contexts (and let you build from the work of others).

The [OGP toolbox](#) is a repository of existing tools working with OCDS data that might be helpful to your project. These tools demonstrate some of the many innovative applications of open contracting data, but should not be thought of as an exhaustive list of all OCDS-based innovations.

Annex I: Global directories of contracting data

I: Public Procurement Data Availability

GLOBAL OPEN DATA INDEX- PROCUREMENT TENDERS		
Data owner	Open Knowledge International	Relevance
Coverage	Global, 2014-2015	Open procurement data
URL	• http://index.okfn.org • http://index.okfn.org/dataset/procurement/	
Description - what information is included in this dataset?	The Global Open Data Index is an annual effort to measure the state of open government data around the world. The crowdsourced survey is designed to assess the openness of specific government datasets according to the Open Definition. The "Procurement Tenders" sections rates the quality of tenders and award data of the national/federal government aggregated by office and provides links to the datasets.	
Relevant history / Related information	The Open Data Barometer and Open Data Index should be merging their data collection in 2017. An updated version may be released during the period of the innovation challenge.	

II: Government Contracts Data Availability

PUBLIC SECTOR OPEN DATA		
Data owner	World Wide Web Foundation	Relevance
Coverage	Global, 2013-2015	Open procurement data
URL	• http://opendatabarometer.org/ • http://opendatabarometer.org/data-explorer/?_year=2015&indicator=ODB.2013.D16&lang=en	
Description - what information is	The Open Data Barometer (ODB) aims to uncover the true prevalence and impact of open data initiatives around the world. It analyses global trends, and provides comparative data on countries and regions using an	

included in this dataset?	in-depth methodology that combines contextual data, technical assessments and secondary indicators. The “Government Contracts” section details of the contracts issued by the national government.
Relevant history / Related information	The Open Data Barometer and Open Data Index should be merging their data collection in 2017. An updated version may be released during the period of the innovation challenge.

III: World Bank Benchmarking Public Procurement

GLOBAL OPEN DATA INDEX- PROCUREMENT TENDERS		
Data owner	World Bank	Relevance
Coverage	Global, 2015-2017	Open procurement data
URL	http://bpp.worldbank.org/ http://bpp.worldbank.org/~media/WBG/BPP/Documents/Reports/Benchmarking-Public-Procurement-2017.pdf	
Description - what information is included in this dataset?	The most recent Benchmarking Public Procurement (BPP) report assesses public procurement regulatory systems in 180 economies. The study reports statistics on the size of public procurement in these economy and provides a cross analysis of regulatory environments that affect the ability of private companies to do business with governments.	