

[General](#)
[The research question](#)
[Research assumptions](#)
[Datasets](#)
[Places](#)
[Scoring](#)
[Conditions](#)
[Sample methodology](#)
[Assessment and quality review process](#)

General

The Global Open Data Index collects and presents information on the current state of open data release around the world. The Global Open Data Index is run by [Open Knowledge International](www.okfn.org) with the assistance of volunteers from the Open Knowledge Network around the world. The first Open Data Index was released on October 28, 2013.

This page explains the methodology behind the Global Open Data Index. If you have any further questions or comments about our methodology please reach out to the staff, community of volunteers, and Index reviewers on the Global Open Data Index Forum.

The Global Open Data Index is not an official government representation of the open data offering in each country, but an independent assessment from a citizen's perspective. It is a civil society audit of open data, and it enables government progress on open data by giving them a measurement tool and a baseline for discussion and analysis of the open data ecosystem in their country and internationally from a key user's perspective.

The Global Open Data Index is not only a benchmarking tool, it also plays a powerful role in sustaining momentum for open data around the world - and in convening civil society networks to use and collaborate around this data. If, for example, the government of a country does publish an open dataset, but this is not clear to the public and cannot be found through a simple search, then the data can easily be overlooked and not put to good use. Governments and open data practitioners can review the Index results to see how accessible the open data they publish actually appears to their citizens, see where improvements are necessary to make open data truly open and useful, and track their progress year to year.

We would like to acknowledge the people who worked on the Global Open Data Index.

****Research team****: Katelyn Rogers and Mor Rubinstei

****Local research coordinators****: Tarek Amr, Paula Alzualde, Oludotun Babayemi, Neal Bastek, Yamila Garcia, Bruce Hoo Fung, Hazwany Jamaluddin, Codrina Maria Ilie, Joachim Mangilima, Matthew McNaughton, Iris Palma.

****Thematic reviewers**** - Tryggvi Bjorgvinsson, Zach Christensen, Stephen gates, Kamil Gregor, Codrina Maria Ilie, Georg Neumann, Yaron Michl, Rebecca Sentance, Gil Zertzer.

****Advisors**** - Jonathan Gray, Rufus Pollock and the community of the Global Open Data Index on the discuss forum.

The research question

Like any other benchmarking tool, the Global Open Data Index tries to answer a question. In our case, the question is as follows:

“What is the state open data around the world?”

From this question, other important questions emerge, such as:

- “Which country ranks best on open data? Who is the least/most open country?”
- “What is the most open dataset? What is the least open dataset?”
- Open data has two key aspects: legal and technical openness. Which of these two — and which specific requirements e.g. an open license, machine readability, bulk access — is the most challenging for data publishers? For example, do governments find it easy to publish machine readable data but struggle to apply an open license?

According to [the common open data assessment framework](<http://opendataresearch.org/sites/default/files/posts/Common%20Assessment%20Workshop%20Report.pdf>) there are four different ways to evaluate data openness — context, data, use and impact. The Global Open Data Index is intentionally narrowly focused on the data aspect, hence, limiting its inquiry only to the datasets *publication* by national governments. It does not look at the broader societal context — for example the legal or policy framework, (FOI, etc.) — and it also does not seek to assess use or impact in a systematic way. Lastly, it does not assess the quality of the data. This narrow focus of data publication enables it to provide a standardized, robust, comparable assessment of the state of the publication of key data by governments around the world.

Research assumptions

Different countries have different governance structures (Federal vs. National government, etc.) and different policies regarding open data. We set out here our key assumptions that inform our approach and that were taken into consideration while collecting and assessing the data.

Assumption 1: Open Data is defined by the Open Definition

We define open data according to the [‘Open Definition’] (<http://opendefinition.org/>)— The open definition is a set of principles that define openness in relation to data and content. It is the, original, “gold-standard” definition for open data. It is also simple and easy to operationalise.

We note one small deviation from the current v2.1 draft of the Definition. The only part of our methodology that is not aligned with the open definition is

“Open Machine readable” format. We give full score to governments that publish data even in a non *open* machine readable format (Such as XLS), this is because we believe governments still lag behind in publishing data in machine readable format, and that we don’t want to discourage them from publishing while they catch up.

Assumption 2: The role of government in publishing data

In the past, there have been questions in the index community about the role of the government in ensuring the publication of a specific dataset. In many fields, some of government services are privatized, which means the data is owned and produced by a company and not the state. For example, in some countries, the drinking water system is run by private companies, and therefore the government is not the data producer and it has to acquire the data from these companies.

Our view and assumption is that __for the key datasets we survey, the national government has a responsibility to ensure the open publication of such data even if it is held and managed by a third-party.__ Therefore, even if the data is not produced by the government, we see it as responsible to ensure the open publication of the data.

Assumption 3: National government as aggregator of data

Not all countries have the same governance structure and have differing degrees of centralisation of services. Some have a main government with municipalities, other have much more complicated structures with sub governments (regions and states). Different governments may collect different data for different geographical regions. It is possible that not all of the sub governments have to abide by the same laws, since they have some autonomy.

__Our assumption is that Federal (or national) government is accountable for the open publication by all its sub-governments.__

In addition, whilst not strictly required, we expect that national governments also provide aggregation of that data from sub governments so as to ensure users have an easy way to access use the data (the best solution is one consolidated dataset but at a minimum could consist of a single point of access to all data subsets).

Datasets

Dataset definitions are crucial in enabling respondents to accurately assess datasets and to do in a way that is comparable across countries. Each year we have refined our definitions and this has continued this year. In addition, this year we have refined the specific guidelines for each dataset:

- **Describe the dataset by at least 3 key data characteristics it must have.**

Currently, we describe each dataset by the key information we want.

For example: Elections results have three qualities (“Results by constituency / district for all major national electoral contests”).

1. The results
2. Geographical data
3. Candidate data

Maps on the other hand used to have only one characteristic: “high level map at a scale of 1:250,000 or better (1cm = 2.5km)”. In order to make this less vague we will consider adding more characteristic in the future, such as marks of public places and borders.

- **Avoid specific indicators, but don’t be afraid to specify if needed.** Some datasets need to include specific data such as names of emissions gases or statistical indicators. In these cases, we tried not to use the condition ‘or’ since it will create a whole new different unit of analysis. Instead, we tried specific. It will help governments to understand what they are missing in their data as well.
- **Include how often the dataset needs to be updated.** Currently, we use the “Is this timely” question in the Index survey. However, different datasets reasonably have different times in which they are updated. Adding this characteristic to the dataset definition can help users answering this question.
- **Aggregation.** Mention which aggregation level the data needs to be in. Some datasets can be in more than one aggregation level and mentioning the aggregation level can help to avoid confusion between datasets.

This year, we added 4 new datasets to the Index. This is the full list of datasets

for 2015:

Name of dataset	Status	Description
National Statistics	pre-existing	Key national statistics such as demographic and economic indicators (GDP, unemployment, population, etc). To satisfy this category, the following minimum criteria must be met: - GDP for the whole country updated at least quarterly - Unemployment statistics updated at least monthly - Population updated at least once a year
Government Budget	pre-existing	National government budget at a high level. This category is looking at budgets, or the planned government expenditure for the upcoming year, and not the actual expenditure. To satisfy this category, the following minimum criteria must be met: - Planned budget divided by government department and sub-department - Updated once a year. - The budget should include descriptions regarding the different budget sections.
Government Spending	pre-existing	Records of actual (past) national government spending at a detailed transactional level. A database of contracts awarded or similar will <i>*not*</i> be considered sufficient. This data category refers to detailed ongoing data on <i>*actual*</i> expenditure. Data submitted in this category should meet the following minimum criteria: - Individual record of transactions - Date of the transactions - Government office which had the transaction - Name of vendor - Amount of the transaction - Updated on a monthly basis
Legislation	pre-existing	This data category requires all national laws and statutes available to be available online, although it is not a requirement that information on legislative behaviour e.g. voting records is available. To satisfy this category, the following minimum criteria must be met: - Content of the law / statutes - If applicable, all relevant amendments to the law - Date of last amendments - Data should be updated at least quarterly
Election Results	pre-existing	This data category requires results by

		constituency / district for all major national electoral contests. To satisfy this category, the following minimum criteria must be met: - Result for all major electoral contests - Number of registered votes - Number of invalid votes - Number of spoiled ballots - All data should be reported at the level of the polling station
National Map	pre-existing	This data category requires a high level national map. To satisfy this category, the following minimum criteria must be met: - Scale of 1:250,000 (1 cm = 2.5km). - Markings of national roads - National borders - Marking of streams, rivers, lakes, mountains. - Updated at least once a year.
Pollutant Emissions	pre-existing	Aggregate data about the emission of air pollutants, especially those potentially harmful to human health (although it is not a requirement to include information on greenhouse gas emissions). Aggregate means national-level or available for at least three major cities. In order to satisfy the minimum requirements for this category, data must be available for the following pollutants and meet the following minimum criteria: - Particulate matter (PM) Levels - Sulphur oxides (SOx) - Nitrogen oxides (NOx) - Volatile organic compounds (VOCs) - Carbon monoxide (CO) - Updated at least once a week. - Measured either at a national level by regions or at least in 3 big cities.
Company Register	pre-existing	List of registered (limited liability) companies. The submissions in this data category do not need to include detailed financial data such as balance sheet, etc. To satisfy this category, the following minimum criteria must be met: - Name of company - Unique identifier of the company - Company address - Updated at least once a month
Location datasets	pre-existing	A database of postcodes/zipcodes and the corresponding spatial locations in terms of a latitude and a longitude (or similar coordinates in an openly published national coordinate system). If a postcode/zipcode system does not exist in the country, please submit a dataset of administrative borders. Data submitted in this category must satisfy the following minimum conditions:

		- Zipcodes * Address * Coordinate (latitude longitude) * national level * updated once a year - Administrative boundaries * Borders poligone * name of poligone (city, neighborhood) * national level * updated once a year
Government procurement tenders (past and present)	new	All tenders and awards of the national/federal government aggregated by office. Monitoring tenders can help new groups to participate in tenders and increase government compliance. Data submitted in this category must be aggregated by office, updated at least monthly & satisfy the following minimum criteria: - Tenders * Tenders name * Tender description * Tender status - Awards * Award title * Award description * value of the award * Suppliers name
Water Quality	new	Data, measured at the water source, on the quality of water is essential for both the delivery of services and the prevention of diseases. In order to satisfy the minimum requirements for this category, data should be available on level of the following chemicals by water source and be updated at least weekly: - fecal coliform - arsenic - fluoride levels - nitrates - TDS (Total dissolved solids)
Weather forecast	new	5 days forecast of temperature, precipitation and wind as well as recorded data for temperature, wind and precipitation for the past year. In order to satisfy the minimum requirements for this category, data submitted should meet the following criteria:

		- 5 days forecast of temperature updated daily - 5 days forecast of wind updated daily - 5 days forecast of precipitation updated daily - Historical temperature data for the past year
Land Ownership	new	Cadaster showing land ownership data on a map and include all metadata on the land. Cadaster data submitted in this category must include the following characteristics: - Land borders - Land owners name - Land size - National level - Be updated yearly

In addition to the 13 datasets above, we have collected data on two more datasets: Transport Timetables and Health performance. The table below describes what we looked for and why we omitted them from the final scoring. Data on the datasets can be found [here](<http://global.census.okfn.org/api/entries.csv>) for direct download.

Name of dataset	Status	Description	Justification for omitting the data
Transport Timetables	pre-existing	Timetables of major government operated (or commissioned) *national-level* public transport services (specifically bus and train). The focus here is on national level services (not those which operate *only* at a municipal or city level and which are not controlled or regulated by the national government). A 'yes' in any question will refer to both types of transport. However, if there is no national level service operated or regulated by the government for a given type of transport (for instance busses), then this type is ignored in this data category. Data submitted in this category should meet the following minimum criteria: - Time of operating - Time of leaving first station and arriving to the last station - Updated at least once a year	During the review of the data, the review team have found out that there 45 places do know if the government collect the data or what is the common mean of transport in between cities in the country. Since there is a confusion around this data, we decided to omit it and do devote more time next year in order to see how we can incorporate it in the global index. In addition, different local index still collect the data on a local level.
Health	new	Geo location of public hospitals and	In this dataset, we asked

Performance		health facilities with opening hours and infectious diseases rate, updated at least once a year. Data submitted in this category must include the following: - Location of public hospitals and clinics. - Data on infectious diseases rates in a country.	for two different datasets, but didn't give room in the questionnaire to answer both datasets. Therefore, we could not evaluate these dataset and score them properly. We did save all of the data, and we will analyse it in order to define this dataset better for next year.
-------------	--	--	---

Places

In a few cases, we have received submissions for 2013 and 2014 from places that are not officially recognised as independent countries; we have included these if they are complete and accurate submissions. Therefore, the Global Open Data Index 2015 ranks 'Places' and not 'Countries'.

Generally we seek to survey jurisdictions with sufficient autonomy to be responsible for data management and publication. Usually these are countries, however there are cases where country jurisdiction is disputed and we generally seek to be flexible and inclusive where we can.

Scoring

Each dataset in each place is evaluated using nine questions that examine the openness of the datasets based to the open definition.

The following table describes the questions in further details alongside their weights.

- High weights were given to questions we assessed to be critical in opening data
- 30 points were given to the open license question, a topic which is still problematic in open data implementation and re-use.

- Open license is the key aspect of actually being open data. Many people release some data but license it in a way that creates a barrier to use, and especially restricts reuse which is critical.
- 15 points were given to the machine readable question, since without data being machine readable, it is hard to reuse and reveal the potential of the data.
 - Again, a major barrier and another area where governments frequently fall down. Reuse is significantly impacted without machine readability and a lack of machine readability can rob data that is openly licensed of any actual practical “openness.”
- 15 points were given for data being free of charge since because, again, it presents a major obstacle in terms of being practically open.

Final Scoring -

After all data is submitted and reviewed, countries are ranked according to their percentage of openness. The percentage is calculated by adding all of the datasets scores (see details on scoring in the table below) and divide them by 1300 (the maximum possible score that a country can get) -

$\text{sum}(13 \text{ dataset}) / 1300 = \text{Index percentage}$

The percentages are rounded to the nearest whole number.

Full scores of datasets can be shown in the places page for each country or in the data dump. You can download the raw Index data

[here](<http://2015.index.okfn.org/download/>)

The following table describes the questions and their scoring weights:

Question	Details	Weighting
Does the data exist?	Does the data exist at all? The data can be in any form (paper or digital, offline or online etc). If it is not, then all the other questions are not answered.	5

Is the data in digital form?	This question addresses whether the data is in digital form (stored on computers or digital storage) or if it only in e.g. paper form.	5
Publicly available?	This question addresses whether the data is "public." This does not require it to be freely available, but does require that someone outside of the government can access it in some form (examples include if the data is available for purchase, if it exists as a PDF on a website that you can access, if you can get it in paper form — then it is public). If a freedom of information request or similar is needed to access the data, it is not considered publicly available.	5
Is the data available for free?	This question addresses whether the data is available for free or if there is a monetary cost to obtain it. If there is a cost, then that is stated in the comments section.	15
Is the data available online?	This question addresses whether the data is available online from an official source. In cases where this is answered with a 'yes', then the link to the data is captured.	5
Is the data machine-readable?	Data is machine-readable if it is in a format that can be easily analysed by a computer. Data can be digital but not machine-readable. For example, consider a PDF document containing tables of data. PDFs are definitely a digital format, but they are not considered machine-readable because a computer would struggle to access the tabular information. The equivalent tables in a format such as a spreadsheet would be considered machine-readable. Note: The appropriate machine-readable format may vary by type of data —	15

	so, for example, machine-readable formats for geographic data may be different than for tabular data. In general, HTML and PDF are not machine-readable formats	
Available in bulk?	Data is available in bulk if the whole dataset can be downloaded or accessed easily and at once. Conversely it is considered non-bulk if citizens are limited to just getting parts of the dataset (for example, if one is restricted to querying a web form and retrieving a small sample of results at a time from a very large database).	10
Openly licensed?	This question addresses whether the dataset is open as per The Open Definition (http://opendefinition.org). To be considered open, the licence or the terms of use needs to explicitly state that anyone is allowed to freely use, reuse or redistribute the data (subject at most, to attribution or share alike requirements). It is vital that a licence is available in some form (if there is no licence, the data is not openly licensed). Open licences that meet the requirements of the Open Definition are listed at http://opendefinition.org/licenses/ .	30
Is the data provided on a timely and up to date basis?	This question addresses whether the data is up to date and timely – or long delayed or outdated. For example, is election data made available immediately or soon after an election or is only available many years later? Timeliness of data is a defined attribute of the dataset characteristics.	10

Conditions

In order to increase validity, we have set conditions to the some of the question. See full conditions list in [this survey flow](<http://census.okfn.org/survey-flow/>)

Sample methodology

The Index uses a [non-probability sampling technique](<http://sociology.about.com/od/Research/a/sampling-designs.htm>) – also known as a “snowball sample”. A snowball sample tries to locate subject of studies in areas that are hard to locate. In our case, we work with contributors who are interested in open government data activity who can assess the availability and quality of open datasets in their respective locations. We do so not only by using referrals, but also by reaching out on social media, through regular communications our Open Government Data and [Open Data Index forums](<https://discuss.okfn.org/c/open-data-index/global-open-data-index-2015>), and by actively networking at conferences and events. This year, we also hired local coordinators, that outreached to their networks and assist in soliciting new submissions. This means that anyone from any place can participate and contribute to the Global Open Data Index as a contributor and make submissions, which are then reviewed. We do not have a quota on the number of places that can participate. Rather, we aim to sample as many places around the world as we can.

This also has an impact on the quality of the data we collected in the first stage of the Global Open Data Index. Contributors have diverse knowledge and backgrounds in open data and therefore they sometimes need help finding the data we are looking for. The following section explains how we tried to deal with this problem.

In 2014 we permitted the “bringing forward” of data submitted in the previous year with only basic review. This meant that some countries which did not have new data in 2014 but did submit in 2013 were brought forward as is to the 2014 census. Although this was mentioned in the Index methodology, in some minor cases their ranking still raised questions. This year, we will require review of all

data and data will not be “brought forward” unless it has been reviewed again and found to be satisfactory. This practice was not done in 2015. Only countries that submitted all datasets for 2015 were included.

Assessment and quality review process

The 2015 assessment and review of the datasets took place in four steps. The first step is collecting the evaluation of datasets through volunteer contributors, and the second step is QA checks by the local coordinators, the third is verifying the results with paid expert reviewers. The fourth and the last step is a public review of the Index before it is published. The following steps are taken each time a dataset is submitted:

1. Contributors, which can be any person, submit information about the availability of one of the key datasets in their Place. At this stage, the results are not published online straight away. Instead their contribution is held for review (see below). Please note that in some Places, when there were no changes in a dataset between years (for example, from 2013 to 2014), the community coordinators of the Index may transfer the existing entry over to the next year, without further ado, while adding in the comments section of that entry that there were ‘no changes from year X to year Y’.
2. Next, our community coordinator does a quick QA check over the data to make sure it doesn’t have any mistake or missing values.
3. Once the review process is done, the entries are then scrutinized by a panel of expert reviewers, who will review submission in a thematic manner. This means that each reviewer will review submissions from all countries under the same dataset. This is done to increase validity.
4. Lastly, we open the Index after the thematic review for a final round of public comments. This was done in order to reduce the error that can be done due to language barriers and to reduce false positives. In this stage, new submissions were allowed. Each submission was reviewed by the Index research lead and if needed, by the thematic reviewers.

5. Please note that this extended kind of quality control was not done for the 2013 and 2014 Open Data Index. This may explain some of the differences in scores when comparing this year's results to the 2013 and 2014 results.

For an extended model of the review please [[see this doc](#)].

Each thematic reviewer was free to establish their own review logic. This allowed for better consistency and clarification. Below are further details of the reviewers' logic:

Government Budget

Reviewer: Mor Rubinstein

The stated description of the Government Budget dataset is as follows:

National government budget at a high level. This category is looking at budgets, or the planned government expenditure for the upcoming year, and not the actual expenditure. To satisfy this category, the following minimum criteria must be met: Planned budget divided by government department and sub-department Updated once a year. The budget should include descriptions regarding the different budget sections.

Submissions that included data for both department AND sub-department/program were accepted. Submissions that included only department level data were not accepted. Additionally, budget speeches that did not include detailed data about the the estimated expenditures for the coming year were not accepted as a submission. Only datasets from an official source (e.g. The Ministry of Finance or equivalent agency) were accepted.

Government Spending

Reviewer: Tryggvi Björgvinsson

The stated description of the Government Spending dataset is as follows:

Records of actual (past) national government spending at a detailed transactional level; A database of contracts awarded or similar will not be considered sufficient. This data category refers to detailed ongoing data on actual expenditure. Data submitted in this category should meet the following minimum criteria: Individual record of transactions. Date of the transactions Government office which had the transaction Name of vendor amount of the transaction Update on a monthly basis

Submissions that included aggregate data or simply procurement contracts (results of calls for tenders) were not accepted. In cases where aggregate data or procurement data was submitted or the submitter claimed that the data did not exist, an attempt was made to locate transactional data with a simple Google search and/or via IBP's Open Budget Survey. If data was available for the previous year (or applicable recent budget cycle) the submission was adjusted accordingly and accepted.

Election Results

Reviewer: Kamil Gregor

The stated description of the Election Results dataset is as follows:

This data category requires results by constituency / district for all major national electoral contests. To satisfy this category, the following minimum criteria must be met: Result for all major electoral contests Number of registered votes Number of invalid votes Number of spoiled ballots All data should be reported at the level of the polling station

Submissions that did not show the data at polling station level were omitted and marked as 'Data does not exist', even if votes are not counted at polling station level as a matter of policy. The reason for this is the polling station level is the most granular level that allow to monitor election fraud .

Company Register

Reviewer: Rebecca Sentance

The stated description of the Company Register dataset is as follows:

List of registered (limited liability) companies. The submissions in this data category does not need to include detailed financial data such as balance sheet etc. To satisfy this category, the following minimum criteria must be met: Name of company Unique identifier of the company Company address Updated at least once a month

Data was marked as unsure if it exists when the submitted dataset did not contain address or a company ID. If the submission referenced a relevant government website that does not indicate the data exists, or if there is no evidence even which government body would hold the data, the submission was changed to 'data does not exist'. If it is clear that a governmental body collects company data, but there is no way of knowing what it consists of, where it is held, or how to access it, and no indication that it would fulfil our requirements, the submission was also marked as 'data does not exist'.

Based on the definition, it was decided that a company register that is freely available to searchable by the public but requires entering a search term (search applications) did not count as free or publicly accessible. However, a company register that can be browsed through page-by-page does present all of the data and is the type of dataset required for acceptance.

National Statisticsational laws and statutes available to be available online, although it is not a requirement that information on legislative behaviour e.g. voting records is available. To satisfy this category, the following minimum criteria must be met: Content of the law / status If applicable, all relevant amendments to the law Date of last amendments Data should be updated at least on quarterly

Submissions were reviewed to ensure the data met the criteria. Regularity of updating was assessed based on the date of the most recently submitted data.

Pollutant Emissions

Reviewer: Zach Christensen

The stated description of the National Statistics dataset is as follows:

Key national statistics such as demographic and economic indicators (GDP, unemployment, population, etc). To satisfy this category, the following minimum criteria must be met: GDP for the whole country updated at least quarterly
Unemployment statistics updated at least monthly
Population updated at least once a year

For each submission, the reviewer checked for national accounts, unemployment, and population data as required by the description. It was found that most countries don't have these data for the last year and very few had quarterly GDP figures or monthly unemployment figures. Submissions were only marked as 'data does not exist' if they did not have any national statistics more recent than 2010.

Legislation

Reviewer: Kamil Gregor

The stated description of the Legislation dataset is as follows:

This data category requires all nReviewer: Yaron Michl

The stated description of the Pollutant Emissions dataset is as follows:

Aggregate data about the emission of air pollutants especially those potentially harmful to human health (although it is not a requirement to include information on greenhouse gas emissions). Aggregate means national-level or available for at least three major cities. In order to satisfy the minimum requirements for this category, data must be available for the following pollutants and meet the following minimum criteria: Particulate matter (PM) Levels Sulphur oxides (SOx) Nitrogen oxides (NOx) Volatile organic compounds (VOCs) Carbon monoxide

(CO) Updated on at least once a week. Measured either at a national level by regions or at least in 3 big cities.

VOCs is a generic designation for many organic chemicals, therefore, when measuring VOCs it is possible to measure any one of a number of compounds such as Benzene or MTBE. Measurements of Volatile Organic compounds (VOCs) was ultimately not considered as part of the data requirements because of this discrepancy and the fact that it is rarely measured on a national level (see this link).

Carbon monoxide (CO) and Nitrogen Oxides (NoX) were also not considered as a requirement because their main origin is usually from transportation.

In addition, some countries publish air pollution by using the Air Quality Index, a formula that translates air quality data into numbers and colors to help citizens understand when to take action to protect their health. Submissions that relied on the Air Quality Index was considered not to exist because it is not raw data.

Government Procurement Tenders

Reviewer: Georg Neumann

The stated description of the Government Procurement Tenders dataset is as follows:

All tenders and awards of the national/federal government aggregated by office. Monitoring tenders can help new groups to participate in tenders and increase government compliance. Data submitted in this category must be aggregated by office, updated at least monthly & satisfy the following minimum criteria:

Tenders: tenders name tender description tender status Awards: Award title
Award description value of the award suppliers name

Quality of published information varied strongly and was not evaluated here. As long as the minimum information was available the data was said to exist for a given place.

Thresholds for publication of this information varies strongly by country. For all EU countries, tenders above a specific amount, detailed here, need to be published. This allowed for all EU submissions to qualify as publishing open procurement data even though some countries, such as Germany, do not publish award value for contracts below those thresholds, and others have closed systems to access specific information on contracts awarded.

In other countries not all sectors of government publish tenders and awards data. Submissions were evaluated to ensure that the main government tenders and contracts were made public, notwithstanding that data from certain ministries may have been missing.

Water Quality

Reviewer: Nisha Thompson

The stated description of the Water Quality dataset is as follows:

Data, measured at the water source, on the quality of water is essential for both the delivery of services and the prevention of diseases. In order to satisfy the minimum requirements for this category, data should be available on level of the following chemicals by water source and be updated at least weekly: fecal coliform arsenic fluoride levels nitrates TDS (Total dissolved solids)

If a country treats water or distributes it, then there will be data regarding water quality because all water treatment requires quality checks. Even though water quality is a local responsibility in most countries, very few countries have a completely decentralized system. Usually there is a monitoring role by the central government, either by the Environmental Protection Agency, Ministry of the Environment or the Ministry of Public Health. If there is monitoring role, the data does exist, if monitoring is completely decentralized, like in the UK, the submission was marked as 'does not exist' because there is no aggregation of the data. If data was not available daily or weekly it wasn't considered timely.

In some cases, all the parameters were accounted for except TDS. Even though it is standard, some countries only collect conductivity, which can be used to calculate TDS. In this case, the submission was approved as is.

Land Ownership

Reviewer: Codrina Maria Ilie

The stated description of the Land Ownership dataset is as follows: Cadaster showing land ownership data on a map and include all metadata on the land. Cadaster data submitted in this category must include the following characteristics: Land borders Land owners name Land size national level updated yearly For various reasons, the land owner's name attribute was widely unmet and as such, lack of this data was not considered a factor in evaluating these submissions. As this dataset is subject to well-kept historic records (not always the case), to legislation (which can be fluctuant), to very expensive activities that a government must implement in order to keep data up to date, to the complexity of the data itself (sometimes data that makes a national cadastre is registred in different registries or systems), a first year indexing exercise must not be considered exhaustive.

Weather

Reviewers: Neal Bastek & Stephen Gates

The stated description of the Weather dataset is as follows:

5 days forecast of temperature, precipitation and wind as well as recorded data for temperature, wind and precipitation for the past year. In order to satisfy the minimum requirements for this category, data submitted should meet the following criteria: 5 days forecast of temperature updated daily 5 days forecast of wind updated daily 5 days forecast of precipitation updated daily Historical temperature data for the past year

Based on a general assessment of the submissions, a minimum threshold for claiming the data existed was set at forecast data for today + two days (three

days) with a qualitative allowance made for arid regions substituting humidity data for precipitation data. The threshold for inclusion could also be met with four day forecasts that include temperature and precipitation data, and/or a generic statement using text or descriptive icons about conditions (e.g. windy, stormy, partly cloudy, sunny, fair, etc.).

Location datasets

Reviewer: Codrina Maria Ilie

The stated description of the Location dataset is as follows:

A database of postcodes/zipcodes and the corresponding spatial locations in terms of a latitude and a longitude (or similar coordinates in an openly published national coordinate system). If a postcode/zipcode system does not exist in the country, please submit a dataset of administrative borders. Data submitted in this category must satisfy the following minimum conditions

Zipcodes	Address	Coordinate (latitude longitude)	national level	updated once a year	
Administrative boundaries	Boarders	poligone	name of poligone (city, neighborhood)	national level	updated once a year

In cases in which a country has not adopted a postcode system, the location dataset is considered to be administrative boundaries. The Universal Postal Union – Postal Addressing System was used to identify the structure of a postcode for a given place

[<http://www.upu.int/en/activities/addressing/postal-addressing-systems-in-member-countries.html>]. This tool proved significantly useful when identifying countries that do not use a postcode system.

In situations where countries only had a postcode search service, either by postcode or address, data was said to not exist. If the postcodes were not geocoded, submissions did not meet the Index requirements due to the difficulty of geocoding such a dataset. On the other hand, if the postcode system took into account just the smallest administrative boundary and if that boundary was

officially available, considering the easiness of obtaining the geocoded postcodes number, the data was marked as 'does exist' for that submission.

National Map

Reviewer: Gil Zaretzer

The stated description of the National Map dataset is as follows:

This data category requires a high level national map. To satisfy this category, the following minimum criteria must be met: Scale of 1:250,000 (1 cm = 2.5km). Markings of national roads National borders Marking of streams, rivers, lakes, mountains. Updated at least once a year.

Only submissions from an official source, with original data was considered. A link to Google Maps, which was often provided, does not satisfy the criteria for these submissions.

In cases where there was no link provided in the submission and entries were marked as "unsure" if there was any indication that the data exists but was not available online, i.e. a national mapping service without a website.