



## Open Data Barometer 3rd edition

Research Handbook - v1.02 24th September 2015

Welcome to the Open Data Barometer Handbook and thank you for participating in the 3rd edition of the Open Data Barometer study. This guide introduces the assessment process of the Open Data Barometer, including information on the methodology, a short tutorial on using the research online platform, guidance on the sources you can use and how to cite them, and **detailed question-by-question scoring guidance and thresholds** to be consulted as your work through the research process. **You should read the first section of this handbook in detail before starting your research.**

### Overview

This document acts as a handbook for researchers, and a reference for those seeking to understand the data that has been made available through the study. This should be considered a living document as we pilot [common methods for assessing open data](#). The most updated online copy can be found at <https://goo.gl/DUkFbE>. Your help and feedback is valuable in assisting us to test, verify and develop the research methodology for this project. You can send your feedback on the methodology and handbook to [project-odb@webfoundation.org](mailto:project-odb@webfoundation.org).

The main aim of this handbook is to provide consistent referencing as a way to **minimize individual interpretations** and confusion when identifying the most appropriate choices for each country score on a given expert survey indicator.

Before answering or reviewing Barometer questions it is also important to familiarise yourself with the concept of **open government data** fully. A good resource for this can be found at <http://opengovernmentdata.org/> and further information can be found in the Open Data Handbook <http://opendatahandbook.org/guide/en/>



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

## SECTION 1 - RESEARCH METHODOLOGY

For the expert survey portion of the Open Data Barometer, a lead country researcher performs desk research and consults with key informants to score the indicators for his/her country. Those initial scores are then reviewed by the project management, who may send back a portion of those questions for improvement or correction based on the standards described in this handbook.

When the project management is satisfied with the state of the indicators, a mix of country, government (by means of a self-assessment questionnaire) and functional expert reviewers will be invited to review the indicators and will flag certain questions for additional research, correction or clarification. Project management will check all reviewer feedback to determine which indicators need to be returned to the lead researcher for further strengthening; this back-and-forth process will lead to a final set of well-researched and well-sourced indicators.

**The time period under study is the 12-month period from July 2014 to June 2015** (both included), although keep also in mind that some relevant evidence could be time-independent (e.g. policies or legislation). All questions (with the exception of C8 and sometimes D9) make reference exclusively to **the national (or federal) government** and can take two main forms:

- A. **'To what extent'** questions, with 11 score choices from 0 to 10 with 10, 5 and 0 score choices and above 0, 3, 5 and 8 incremental scoring guidelines containing detailed scoring criteria to guide the researcher in his/her selection of the most appropriate score.
- B. **'Dataset questions'** broken down into 10 yes/no questions, and a number of prompts for additional evidence.

Most questions can be answered drawing upon online sources, published materials and desk research. Tips on sources and searching for information are provided alongside each question. In a number of cases you may wish to interview or consult with open data experts, NGOs, journalists, government officials or any other actors to identify the appropriate answer to a question. In these cases you must explain clearly to these sources:

- That you are undertaking research for the World Wide Web Foundation Open Data Barometer, a multi-country study of open data policy and practice to be published in the following months;
- That any responses they give may be placed in a public open dataset, and that their responses will be included in this dataset;
- That they are under no obligation to respond to your questions, and can withdraw from

the interview at any time;

You should keep a clear written record, ideally with e-mail trail where interviews or consultations were arranged by e-mail, to show that you have asked each source to consent to the use of their responses in the Open Data Barometer.

### A note on maximum and minimum scores

Although the scores are in the range of zero (0) to ten (10), **allocating a score of 10 for an indicator (question) for a country should be very rare**. A score of 10 would imply virtually no room for improvement, which is not likely to be the case in the large majority of cases. Similar due care should be exercised when allocating a score of zero for an indicator. In both cases, **the evidence needs to be very strong in support of scores at the extremes (10 and 0)**.

Please remember that **the scoring range is continuous and allows for all 11 scores to be used in the range of 0-10**. To guide the researcher in a consistent scoring approach, detailed and incremental criteria based on four incremental scoring threshold levels (above 0; above 3; above 5 and above 8) are provided for each question. Note that the criteria are incremental so, **each level incorporates also the criteria of the previous ones**.

### Providing evidence and sources along with justification

For **every question** it is very important that researchers **provide a brief and reasoned argument to explain the score choice as well as details of the sources used to answer that question**, and any additional information requested. This will support reviewers to check the answers given. **You should always provide a justification**, even when scoring a question zero (0). In these cases you must explain how you tried to locate the details requested in the question. For example, describe the searches you tried to find an open data policy or a particular dataset which ended with no evidence of a policy or that dataset being available.

**For each of the questions you should also provide a 'Confidence Level'** of 0%, 25%, 50%, 75%, 95% or 100% for how certain you are of the answer you have given. Space will also be provided for any **additional notes** in the survey tool. Additional notes will be available to the reviewer, but will not be published. You can use this space to note if you have any areas of uncertainty about a question response, or to record notes to yourself for following up this question further.

You **must** provide your justification and sources, and notes in (proper<sup>1</sup>) **English**. The resources linked to can be in other languages, and you should research and search for sources in the

---

<sup>1</sup> Please, use a grammar and spell checker tool before introducing the results into the survey tool if you need so.

official languages of the country you are researching wherever appropriate. These **justifications should be written in clear prose, and impersonal way with numerical footnotes in (semi-circle brackets)** indicating the relevant sources used, with sources listed at the bottom of the text box. Please, **follow the specified format narrowly**. For example:

Example of justification formatting:

A justification might refer to multiple web pages **(1)** to support the claims made. Include dates of access for the web links. At the relevant point in the justification text you can include footnote references, following the indicated standardised format **(2)**. It should be possible for a reviewer to easily locate all the sources you cite, or to understand the evidence you have to support each statement in your justification.

This keeps the justification prose clear, and ensures that all sources are listed in one place at the bottom of the justification box **(3)** starting with a couple of sharp signs (**##**) Sources header.

Use quotes sparingly. Good justifications rarely exceed three to four paragraphs, and can often be shorter and more concise.

Using semi-circle brackets means that the markdown processor used in the survey tool **(4)** will not distort the formatting of your content when displaying it to a reviewer.

**## Sources**

**(1):** [http://en.wikipedia.org/wiki/Web\\_page](http://en.wikipedia.org/wiki/Web_page) Accessed 12th May 2014.

**(2):** University of Chicago Press, 2010. The Chicago Manual of Style, 16th Edition.

**(3):** Skype interview with Hania Farhan, Web Index Team, 10th May 2014.

**(4):** <http://daringfireball.net/projects/markdown/> Accessed 12th May 2014.

For most questions, where a source is not available online to link to, **you can also upload supporting files through the online survey tool.**

When citing laws or regulations...

- Researchers must identify the law (full name, article, etc.) and provide a direct quote or a detailed explanation of the law in the Comments box, as needed.
- Researchers must refer, where appropriate, to case law, specific legal articles or statutes.
- Researchers are to reflect and describe applicable traditional and customary law when necessary.

- Laws cited must be domestic laws rather than international law.
- Please **use this specific format when citing laws or regulations**: Name of Law, Article, Section Number, Year, Hyperlink to law. Example: *Constitution Acts, Part I: Canadian Charter of Rights and Freedoms, Section 6. 1982.* [http://laws.justice.gc.ca/eng/Const/page-11.html#sc:7:s\\_6](http://laws.justice.gc.ca/eng/Const/page-11.html#sc:7:s_6)

#### When citing interviews and other desk research...

- **Score choices should always reflect the project's official reporting period of 2014 July to June 2015.** Sources and score choices that fall outside the official study period by a few weeks are also allowed; beyond that, however, sources and score choices that fall outside the study period should not be included.
- Exercise professional judgment in determining whether opinions of an interviewee are factual and accurate. **We strongly suggest researchers corroborate information obtained in interviews with desk research.**
- **Please use exactly the following format when citing interviews:** Interview sources must include the full name of the interviewee, the name of the interviewee's employer, job title, date and place (city, country) where the interview was conducted. Example: *Jane Doe, Ministry of Justice, Director General, June 22, 2013, Kampala, Uganda.*
- **Anonymity:** When it is not feasible to publish the name of an interviewee (out of justified fear for the interviewee's safety or negative professional ramifications), please include the name of the interviewee in the "Private Discussion" box of the survey. In the "Sources" box, simply state "Anonymous" for the interviewee name while providing as many other details as possible (e.g. Anonymous, Ministry of Finance, government official, June 1, 2012, Cape Town, South Africa). Web Foundation will maintain that confidentiality and no names will be published that are submitted in the Private Discussion box.
- **Please use the following format when citing media articles:** name of author, name of publication, title of article, date published, and a URL/hyperlink (or a digital/PDF attachment if a link is not available or has expired).
- **Please use the following format when citing journal articles, written documents or other third-party desk research:** provide the full citation and whenever possible please provide a URL/hyperlink (or attach an electronic copy if a link is unavailable). Example: *Comstock, E. W. & Butler, J. W., 2000. Access denied: The FCC's Failure to Implement Open access as Required by the Communication Act. Journal of Communications Law and Policy, Winter.*

- **Researcher's' comments are expected to offer a brief and reasoned argument to explain the score choice.** Make reference to concrete examples or details obtained from the research (interviews, reports, etc.). General or vague comments are not acceptable and will be returned to the researcher for improvement.
- **An isolated example is not enough evidence with which to select a score.** Researchers need to consider and assess the overall context (for example, is a recent scandal/incident reflective of the overall and typical situation or rather a one-time outlier?) before selecting the appropriate score.
- **Personal opinions and broad generalizations without specific supporting evidence within the period of study are not acceptable.** Comments such as “media can sometimes be subjective” “reporters write in exchange for gifts” or “the most powerful judges are close to the government” need to be substituted for specific sources that provide real-life, concrete examples.

You can ask your project manager or the research coordinator to review your first attempt at any of the questions if you are unsure of the answers you have given or the way you are providing justifications.

## SECTION 2: QUESTIONS AND STRUCTURE

The 3rd edition of the Open Data Barometer seeks to repeat the analysis from previous editions, with some small modifications and methodological revisions that are focused on three main aspects:

- A government self assessment simplified questionnaire for each of the countries in the study as an additional source of input for the research.
- Two new additional questions [C2 and C3] and other minor adjustments for all questions as first exploration steps towards the assessment of the International Open Data Charter principles [<http://opendatacharter.net/>].
- A more detailed and incremental scoring guidance with comprehensive criteria and scoring thresholds to guide the researcher and improve consistency of the results.

Overall, however, we have sought to maintain certain consistency with the questions used in previous editions. Wider methodological revisions will continue to be explored in future editions. In any case, **please read the question guidance carefully in each instance and in case you have any doubt don't hesitate to contact the project manager.**

Data provided by the previous edition of the Barometer for each country is available as an initial input to the research. The researcher is expected to challenge, extend, update and keep improving the quality of the answers for all Context, Impact and Datasets sections.

### OPEN DATA CONTEXT QUESTIONS

There are many policy areas that can overlap with open data. However, for our context analysis it is important to understand that open data is conceptually **distinct** from:

- **Open government** - whilst an open government policy might mention open data, the two are not identical. Always check for explicit discussion of open data;
- **E-government** - policies to place government services online *might* have an open data element to them, but in many cases they only give citizens access to specific services or small extracts of data, rather than providing full access to machine-readable data;
- **Data sharing** - governments may increase data sharing between departments, but with limits on whom the data is shared with, or who can re-use it. It is only open data when anyone can re-use it without restrictions;
- **Open Access** - open access focuses mainly on access to (academic) documents and publications rather than datasets.



## C1) To what extent is there an active and well-resourced open government data initiative in the country? [ODB.2013.C.INIT]

### Scoring Criteria

10	5	0
<p>There is a strong national open data initiative with significant resources behind it, including dedicated staff and budgets. There is explicit commitment to open data from a senior government figure (e.g. Cabinet minister) and/or parliamentary backing for an open data initiative.</p>	<p>There is a small-scale open data initiative, or an open data initiative has been announced but is not yet resourced. Senior leadership is making commitments to increased government transparency, and/or some commitments to open data are being expressed by a junior minister / single ministry.</p>	<p>There is no evidence of a formal open data initiative, nor any commitment from government to release open data.</p>

### Evidence and scoring thresholds:

**For a score of above 0** to be awarded: there should be evidence of some open data initiative or at least any commitment from government to release open data.

**For a score of above 3** to be awarded: there should be evidence of a national data catalogue providing access to datasets available for re-use. Access to the data could be provided directly on the catalogue or indirectly through pointers to the place where the data is located.

**For a score of above 5** to be awarded: there should be evidence of a small-scale open data initiative, even when not yet completely resourced. Senior leadership is making commitments to increased government transparency, and/or some commitments to open data are being expressed by at least a junior minister / single ministry.

**For a score above 8** to be awarded: there should be evidence of a strong and consolidated national open data initiative with significant resources behind it, including **dedicated staff and allocated budgets**. There is explicit commitment to open data from a senior government figure (e.g. Cabinet minister) and/or parliamentary backing for an open data initiative.

### Scoring Guidance



Look for evidence of **an explicit government commitment to open data**. An open data initiative is a programme by the government to release government data online to the public. It has four main features:

1. The government discloses data or information without request from citizens. This may be according to a release schedule or *ad hoc*.
2. The Internet is the primary means of disclosure (including mobile phone applications);
3. Data is free to access and re-use, e.g. open licenses;
4. Data is in a machine-readable format to enable computer-based reuse, e.g. spreadsheet formats, Application Programming Interface (API).

Significant resources for an open government data initiative include a sufficient budget, personnel and facilities to carry out the mandate of the open data initiative, including technical personnel with appropriate qualifications for dealing with open data issues.

Note that this question is only concerned with initiatives led by the **national government**. Open data initiatives covering the country, but organised by a third party, such as the African Development Bank or another regional organisation should not be counted, although you can detail these in the 'Additional notes' section.

#### Source Guidance

- The existence of government open data action plans, policies or directives;
- Speeches by government leaders about open data;
- Conversations with government and civil society members of the open data movement;
- Conversations with the “civic hacker” community;
- Open Government Partnership country action plans [\[http://www.opengovpartnership.org/countries\]](http://www.opengovpartnership.org/countries) that contain explicit commitments to open data. See also the database of commitments at [\[https://docs.google.com/spreadsheets/d/1ua7HcCbd69HDKqiz7FW2QKr5ExH4cTNmupuVROdBEeU/edit?usp=sharing\]](https://docs.google.com/spreadsheets/d/1ua7HcCbd69HDKqiz7FW2QKr5ExH4cTNmupuVROdBEeU/edit?usp=sharing);
- Reports published by the media, academic and policy journals, and development and multilateral bodies (e.g. World Bank, IFC, OECD, African Development Bank).
- Regional open data communities, such as the ePSI Platform [\[http://www.epsiplatform.eu/\]](http://www.epsiplatform.eu/), the Caribbean open data Institute [\[http://caribbeanopeninstitute.org/content/open-data\]](http://caribbeanopeninstitute.org/content/open-data) or the Africa Open Data [\[https://www.facebook.com/groups/africaopendata/\]](https://www.facebook.com/groups/africaopendata/), Open Data Francophone [\[https://www.linkedin.com/grp/home?gid=8313551\]](https://www.linkedin.com/grp/home?gid=8313551), Open Government in the Arab States [\[https://www.linkedin.com/grp/home?gid=4850889\]](https://www.linkedin.com/grp/home?gid=4850889) and Open MENA [\[http://openmena.net/en/\]](http://openmena.net/en/) networks.

## C2) To what extent is there a well-defined open data policy and/or strategy in the country? [ODB.2015.C.POLI]

### Scoring Criteria

10	5	0
<p>There is an active national open data strategy defined for a period of at least 2 years. The national data policy establishes a general right to reuse by means of an explicit 'open by default' statement and promotes standard licenses or terms of use to be adopted by the public sector bodies without any possible access and re-use restriction more than attribution and share-alike. Open data training programmes for civil servants are available in order to develop their data analysis and reuse skills. The release of data is considered as part of the regular government performance indicators and progress reports are available.</p>	<p>There is a documented national open data policy or strategy that articulates processes, responsibilities, timelines and resources and a national institution or authority is in charge of its execution. There are general guidelines and standards for data publication covering different aspects such as specific datasets to be published, formats to be used, licensing to be applied, etc. Publication of raw machine readable data and adoption of data standards are clearly promoted.</p>	<p>There is no specific official open data strategy, policy or guidelines.</p>

### Evidence and scoring thresholds:

**For a score of above 0** to be awarded: there should be evidence of any official websites, government documents or guidelines referencing global open data practices in the country, although no formal open data policy or strategy may be yet in place.

**For a score of above 3** to be awarded: there should be evidence of at least some national statements or guidelines on the publication of public sector information, even if just as part of the open government and transparency agenda or any other more general information management programme. A common definition of open data may still not be shared across the public sector and non-commercial restrictions or access fees may exist.

**For a score of above 5** to be awarded: there should be evidence of a documented national open data policy or strategy that articulates processes, responsibilities, timelines and resources and a national institution or authority is in charge of its execution. There are general guidelines and standards for data publication covering different aspects such as specific datasets to be

published, formats to be used, licensing to be applied, etc. Publication of raw machine readable data and adoption of data standards are clearly promoted.

**For a score above 8** to be awarded: there should be evidence of an active national open data strategy defined for a period of at least 2 years. The national data policy establishes a general right to reuse by means of an explicit 'open by default' statement and promotes standard licenses or terms of use to be adopted by the public sector bodies without any possible access and re-use restriction more than attribution and share-alike. General open data training programmes for civil servants are available in order to develop their data analysis and reuse skills. The release of data is considered as part of the regular government performance indicators and progress reports are available.

### Scoring Guidance

The ten principles of open government data are:

1. Completeness
2. Primacy
3. Timeliness
4. Ease of Physical and Electronic Access
5. Machine readability
6. Non-discrimination
7. Use of Commonly Owned Standards
8. Licensing
9. Permanence
10. Usage Costs

For more information on what these principles entail see:

[\[http://sunlightfoundation.com/policy/documents/ten-open-data-principles/\]](http://sunlightfoundation.com/policy/documents/ten-open-data-principles/)

Other open data reference policies are those proposed by the G8 open data charter [\[https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex\]](https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex) and the recently announced International open data charter (currently on consultation phase) [\[http://opendatacharter.net/charter/\]](http://opendatacharter.net/charter/), where the same principles have been refined to support broader global adoption:

- Principle 1: Open Data by Default;
- Principle 2: Quality and Quantity;
- Principle 3: Accessible and Usable by All;
- Principle 4: Engagement and empowerment of citizens (data for improved governance);
- Principle 5: Collaboration for development and innovation (data for innovation).

Look for all these features in the policy you are assessing for it to receive the maximum scores.

## Source Guidance

- The existence of government open data action plans, guidelines, strategies, policies or directives;
- The existence of supranational open data policies or directives affecting the country;
- The existence of other public sector information or information innovation programmes;
- Speeches by government leaders about open data;
- Conversations with members of the open data movement, both government and civil society;
- Open Government Partnership country action plans [<http://www.opengovpartnership.org/countries>] that contain explicit commitments to open data. See also the database of commitments at [[https://docs.google.com/spreadsheets/d/1ua7HcCbd69HDKqiz7FW2QKr5ExH4cTNmup\\_uVROdBEEU/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1ua7HcCbd69HDKqiz7FW2QKr5ExH4cTNmup_uVROdBEEU/edit?usp=sharing)];
- Reports published by the media, academic and policy journals, and development and multilateral bodies (e.g. World Bank, IFC, OECD, African Development Bank).

### C3) To what extent is there a consistent (open) data management and publication approach? [ODB.2015.C.MANAG]

#### Scoring Criteria

10	5	0
<p>There is a single and exhaustive data inventory for the central government, including non published data. Complete user's guides and supporting documentation for reference data are available, including information on the purpose of the data, their descriptions and characteristics, provenance and data collection information, etc.</p> <p>There is a quality control process for the data provided through the catalog covering various aspects such as completeness, granularity, timeliness, persistence, etc.</p>	<p>There is a standardized release process for the publication of data sets, addressing also future updates. Comprehensive machine readable metadata is regularly provided. Multiple format options are usually available for each of the published datasets. There is a set of technical standards (including metadata, data models, codelists and identifiers) for the publication of open data.</p>	<p>There is no specific data management approach and data is provided in an inconsistent way.</p>

#### Evidence and scoring thresholds:

**For a score of above 0** to be awarded: there should be at least some minimal description of the datasets through the provision of metadata, even when it may not be normalised and could be used in an inconsistent way.

**For a score of above 3** to be awarded: there should be shared common metadata elements used across government. Regular public consultations on the user's data needs and preferences are conducted (through online systems, social media, workshops, etc.) and requests are being addressed and responded.

**For a score of above 5** to be awarded: there should be a standardized release process for the publication of data sets, addressing also future updates. Comprehensive machine readable metadata is regularly provided. Multiple format options are usually available for each of the published datasets. There is a set of technical standards (including metadata, data models, codelists and identifiers) for the publication of open data.

**For a score above 8** to be awarded: there should be a single and exhaustive data inventory for the central government, including non published data. Complete user's guides and supporting documentation for reference data are available, including information such as the purpose of the data, their descriptions and characteristics, provenance and data collection information, etc. There is a quality control process for the data provided through the catalog covering various aspects such as completeness, granularity, timeliness, persistence, etc.

### Scoring Guidance

When releasing data, one should aim to do so in a way that helps people to use and understand them. Data needs to be fully described, as appropriate, to help users to fully understand the data. Following the recommendations of the G8 Technical Annex [<https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex#technical-annex>] this may include:

- Documentation that provides explanations about the data fields used;
- Data dictionaries to link different data; and
- A user's guide that describes the purpose of the collection, the target audience, the characteristics of the sample, and the method of data collection.

At the same time, governments need to listen to feedback from data users to improve the breadth, quality and accessibility of data they offer. This could be in the form of a public consultation, discussions with civil society, creation of a feedback mechanism on the data portal, or through other appropriate mechanisms.

### Source Guidance

- The existence of government data management guidelines and data standards policies;
- Documentation of official open data catalogs;
- Evidence of adoption of international reference metadata and data standards (e.g. [DCAT](#), [DCAT-AP](#), [oData](#), [Best Practices for the Publication of Data on the Web](#), etc.);
- Conversations with members of the open data movement, both government and civil society;
- Conversations with government officials working in open data offices or projects;
- Conversations with the "civic hacker" community;
- Conversations with data entrepreneurs and researchers;
- Reports published by the media, academic and policy journals, and development and multilateral bodies (e.g. World Bank, IFC, OECD, African Development Bank).

## C4) To what extent is there a robust legal or regulatory framework for protection of personal data in the country? [ODB.2013.C.DPL]

### Scoring Criteria

10	5	0
A legal or regulatory data protection framework exists that is broadly applicable, provides the right of choice/consent to individuals, provides the right to access and/or correct one's personal data, imposes clear responsibilities on information holders and provides a right of redress against both private and public bodies that violate data privacy.	A legal or regulatory regime exists but is missing some of the key elements understood to promote best practice around data protection policies, including broad applicability, the right of choice/consent to individuals, the right to access and/or correct one's personal data, clear responsibilities on information holders, and/or the right of redress against both private and public bodies that violate data privacy.	A legal or regulatory regime to promote data protection does not exist or is so devoid of precision and/or the understood best practice as to render it useless in practice.

### Evidence and scoring thresholds:

**For a score of above 0** to be awarded: there should be evidence of a legal or regulatory policy to promote data protection of some form.

**For a score of above 3** to be awarded: there should be evidence that the legal or regulatory policy that promotes data protection is concise and useful in practice. some cases where it was used and applied should be provided.

**For a score of above 5** to be awarded: there should be evidence that the legal or regulatory policy that promotes data protection exists but is missing some of the key elements understood to promote best practice around data protection policies, including broad applicability, the right of choice/consent to individuals, the right to access and/or correct one's personal data, clear responsibilities on information holders (data-controllers), and/or the right of redress against both private and public bodies that violate data privacy.

**For a score above 8** to be awarded: there should be evidence that the legal or regulatory law that promotes data protection exists and is applicable, provides the right of choice/consent to individuals, provides the right to access and/or correct one's personal data, imposes clear responsibilities on information holders (data-controllers) and provides a right of redress against



both private and public bodies that violate data privacy.

## Scoring Guidance

Strong data protection regimes include the key features noted below:

(1) **Broad applicability** – these rules should apply to personal data sets and data controllers in both the public and private sectors.

(2) **The right of choice/consent** – Individuals should normally be given the choice of whether their information is collected. There should be only limited exceptions to this where there is an overriding interest, defined in law, in the collection of such information. This implies that individuals understand and are given clear notice of a public or private body's information practices before any personal information is collected. This notification should describe what information is proposed to be collected and held, who will collect it, how the information will be used and who will have access to it. It should also be clear to the subject whether the provision of the requested information is voluntary or required by law and of the consequences of refusing to provide the requested information. Information should not be used for purposes that are incompatible with the use for which the information was originally collected.

(3) **The right to access and correct** – Individuals should have the right of access to any information held about them at reasonable intervals and without undue delay. They should also have the right to require the data controller to correct any inaccuracies or to delete the data, where appropriate.

(4) **The responsibilities of information holders** – Data controllers must take reasonable steps to ensure that the information they hold is accurate and secure. Access to the data should be limited in accordance with the established uses of the data. Transfers should be made only to third parties that can ensure similar respect for data protection principles. Data should be destroyed once the information is no longer needed for the established uses, or converted to anonymous form. While information is held, appropriate steps should be taken to ensure the confidentiality, integrity and quality of the data.

(5) **The right of redress** – Individuals should have the right of redress against public and private bodies that fail to respect data protection rules in relation to data about them. Remedies can be provided through self-regulation, private law actions and government enforcement. Oversight of the system should be undertaken by an independent body.

## Source Guidance

- This 2012 paper [[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2000034](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2000034)] provides a list of 89 countries with data protection laws and the names of the laws to assist in researching whether a law is effectively implemented;
- National and supranational data protection legislation;
- Interviews with senior officials from the information commission or data protection commission;
- Interviews with NGO officers with expertise in data protection and access-to-information issues and also investigative journalists;
- Experts in data privacy, information privacy or access to information, such as academics, researchers and think tanks;
- News and reports published by the media, academic and policy journals, and development and multilateral bodies (e.g., World Bank, IFC, OECD, African Development Bank).

## C5) To what extent does the country have a functioning Right to Information (RTI) / Freedom of Information (FoI) law? [ODB.2013.C.RTI]

### Scoring Criteria

10	5	0
<p>Citizens generally receive responses to requests for government information within a reasonable time period and at a reasonable cost, and responses are typically of acceptable quality. An RTI or FOI law or similar legal right enshrines the right to such requests. A redress mechanism exists and it is effective. Proactive publication is contemplated with requirements on formats that enable reuse.</p>	<p>Citizens may not receive timely responses to requests for government information, or responses are missing key information. The right to request government information may be generally guaranteed by a vague constitutional right but is not further protected by a dedicated law or enforceable regulation.</p>	<p>Requests for government information are generally not responded to at all, or responses are of extremely low quality. There may be no right-to-information or FOI law at all.</p>

### Evidence and scoring thresholds:

**For a score of above 0** to be awarded: there should be evidence of some form of legal or regulatory right to information from government, even though it might not be implemented. Policy statements have been made that this issue will be addressed in government policy, although implementation remains weak.

**For a score of above 3** to be awarded: there should be evidence that the legal or regulatory right to information law is responded to when there are requests for the information, through a dedicated agency or channel, although the response time may be slow and the quality of the information provided may not be as requested. Implementation is patchy and not yet widespread.

**For a score of above 5** to be awarded: there should be evidence that the legal or regulatory regime which exists guarantees citizens access to information. A dedicated agency exists to deal with enquiries and to adjudicate cases or request for information from government that are refused. Response rate from this agency is fairly prompt (within a few months).

**For a score above 8** to be awarded: there should be evidence that citizens generally receive responses to requests for government information within the legally stipulated time as governed by the RTI / FOI law and at the cost defined by law. The responses are typically of acceptable quality. An RTI law or similar legal right enshrines the right to such requests. There is a redress

mechanism with a dedicated agency that adjudicates cases that are refused by the government, and this role is taken seriously and there is evidence of its work being effective and respected. The law also foresees a **right to data** with the obligation to publish data proactively as **open by default** (with some reasonable security and privacy exceptions) and sets requirements on formats for the disclosure of information that enable a general right to reuse.

### Scoring Guidance

This indicator addresses whether the country's disclosure requirements are "effective." The basic requirements for them to be considered "effective" are whether information is:

- A. available to the public for free or at reasonable/minimal costs in a variety of venues (e.g., online, government agency offices);
- B. can be accessed by citizens within a timeframe as defined by the law; and
- C. answers the specific request, with explanations for refusal to release information.

For a 10 score, there can be exceptions in which information is not released to protect national security or public interests clearly prescribed by law (e.g., medical records, sexual orientation, etc.), but the legal reason must be stated clearly in the response from the government to the citizen who requested the information. Proactive publication with reusable formats is also required.

### Source Guidance

- Databases of right-to-information laws: RTI Rating [[www.rti-rating.org](http://www.rti-rating.org)] and Public Accountability Mechanisms [[www.agidata.org/pam](http://www.agidata.org/pam)];
- Assessment of RTI laws in practice: Global Integrity, Category I-3 [[www.globalintegrity.org](http://www.globalintegrity.org)];
- Interviews with NGO officers with expertise in access-to-information issues and investigative journalists;
- If there is a national agency in charge of handling appeals to denials for request of information, then researchers should interview government officials who work for this agency to obtain a sense of the conditions in which an agency denies (or grants) information;
- Reports published by the information agency, media reports and publications by development/donor agencies;
- For researchers to find statistics, they should refer to the national statistical agency or the appropriate governmental agency that houses statistical information;
- Non-exhaustive databases of right-to-information laws can be found at RTI Rating [[www.rti-rating.org](http://www.rti-rating.org)] and from the Public Accountability Mechanisms [[www.agidata.org/pam](http://www.agidata.org/pam)] database;
- You can check existing assessments of RTI laws in practice: Global Integrity, Category I-3 [[www.globalintegrity.org](http://www.globalintegrity.org)];

- Conversations with NGO officers with expertise in access-to-information issues and investigative journalists;
- Reports published by the information agency, media reports and publications by development/donor agencies;
- Principles on national security and right to information  
[\[http://www.opensocietyfoundations.org/sites/default/files/global-principles-national-security-10232013.pdf\]](http://www.opensocietyfoundations.org/sites/default/files/global-principles-national-security-10232013.pdf)

## C6) To what extent are civil society and information technology professionals engaging with the government regarding open data? [ODB.2013.C.CSOC]

### Scoring Criteria

10	5	0
<p>There are coordinated campaign(s) calling for open data or working with governments to promote open data and create value. The government regularly engages with the user community and there are regular outreach events and activities to promote the use of open data, including at least one regular national or inter-regional event. Non governmental organizations are contributing to the enrichment of government data and public-private partnerships are being explored.</p>	<p>There is some demand or engagement over open data from individuals or communities, but it may be isolated. There is no coordinated campaign for open data usage, and the government does not engage the community regularly.</p>	<p>There is little, if any, evidence of demand or use of open data.</p>

### Evidence and scoring thresholds:

**For a score of above 0** to be awarded: there should be evidence of some minimal demand or use of open data by civil society and/or information technology professionals.

**For a score of above 3** to be awarded: there should be evidence of some data users' engagement and interaction, even when it may be basically just through the use of online services to interact, communicate and reach out to the community (e.g. blogs, social media, etc.)

**For a score of above 5** to be awarded: there should be evidence of demand or engagement over open data from individuals or communities, although it may be isolated and with no coordinated campaign for open data usage. At this stage the government does not engage with the community regularly although it shows proactive efforts towards it, such as specific public consultations.

**For a score above 8** to be awarded: there should be evidence of coordinated campaign(s) calling for open data or working with governments to promote open data and create value. The government regularly engages with the user community and there are regular outreach events and activities to promote the use of open data, including at least one regular national or

inter-regional event with government participation that serves as a forum for the discussion of the status of open data in the country. Non governmental organizations are contributing to the enrichment of government data and public-private partnerships are being explored.

### Scoring Guidance

Campaigns for open data are often composed of civil society organizations, data technologists, informational professionals, computer experts and ordinary citizens who advocate for greater access to government data. A 10 score indicates that government officials recognize these organized campaigns and engage regularly in discussion with community leaders about which data to release, when and in what forms, exploring also collaborations.

### Source Guidance

- Online evidence of an open data community, including reports of events and other activities;
- Conversations with open data specialists in civil society organizations or individuals who are directing open data campaigns;
- Conversations with government officials working in open data offices or projects;
- Conversations with NGO officers with expertise in open-data and access-to-information issues, and investigative journalists;
- Reports published by the media, academic and policy journals, and development and multilateral bodies (e.g., Open Knowledge Foundation, Sunlight Foundation, etc.)
- Regional open data communities, such as the ePSI Platform [<http://www.epsiplatform.eu/>], the Caribbean open data Institute [<http://caribbeanopeninstitute.org/content/open-data>] or the Africa Open Data [<https://www.facebook.com/groups/africaopendata/>], Open Data Francophone [<https://www.linkedin.com/grp/home?gid=8313551>], Open Government in the Arab States [<https://www.linkedin.com/grp/home?gid=4850889>] and Open MENA [<http://openmena.net/en/>] networks.



## C7) To what extent is government directly supporting a culture of innovation with open data through competitions, grants or other support actions? [ODB.2013.C.SUPIN]

### Scoring Criteria

10	5	0
<p>The government has made a substantial continuous commitment to support a culture of innovation with open data, including significant financial incentives and supporting a range of different activities. These include events targeting private sector use of open data and sectoral approaches. Multiple government departments are involved in supporting innovation with open data and there are dedicated teams in government charged with exploring data possibilities.</p>	<p>There are a range of interventions to support a culture of innovation. For example, there are three or more examples of competitions, funding schemes or hackathon events run by government AND two or more departments or agencies are involved in running these schemes.</p>	<p>There is no government support for innovation using open data.</p>

### Evidence and scoring thresholds:

**For a score of above 0** to be awarded: there should be some isolated efforts to support a culture of innovation, but these actions may come exclusively from the civil society side with no government support.

**For a score of above 3** to be awarded: there should be a range of interventions to support a culture of innovation, including challenges, hackathons, informative sessions or co-creation sessions, but government is not directly involved and its support may be only testimonial.

**For a score of above 5** to be awarded: there should be evidence of a range of frequent interventions to support a culture of innovation. For example, there are three or more examples of competitions, funding schemes or hackathon events within the year with the participation of the government and two or more different departments or agencies are involved in running these schemes.

**For a score above 8** to be awarded: there should be evidence of substantial continuous commitment from government to support a data-driven culture of innovation with open data, including significant financial incentives to create new services and value and supporting a range of different activities such as the promotion of data entrepreneurship. These include events targeting private sector use of open data and sectoral approaches. Multiple government departments are involved in supporting innovation with open data and there are dedicated

teams in government charged with exploring data possibilities and developing new products and tools.

### Scoring Guidance

Governments can adopt a range of approaches to stimulate a culture of innovation around open data including:

- **Running competitions** in which prize money is offered to innovators creating tools, services or commercial applications using open data;
- **Organising [hackathon](#) events** which invite developers to create prototype tools and services over one or two day events;
- **Organising incubators and open data boot camps** specifically targeted at supporting innovative uses of open data;
- **Offering grant funding or [innovation vouchers](#)** specifically targeted at encouraging businesses to engage with open data;

Sometimes these are run as one-off activities, organised with minimal budgets by small groups of staff. In these cases, a maximum score of 5 should be given. To receive a score of 10 there should be clear evidence of government dedicating investment to support innovation with open data, including support to private sector re-users of open data.

If you find a number of non-government activities to support innovation with open data you may include details of these in the additional notes field, but only events with significant government involvement should feature in the score you give.

### Source Guidance

- Conversations with open data activists, government officials involved in open data, or entrepreneurs working on open data in the country;
- Search for government press releases and announcements about support for open data innovation activities;
- Search for donor press releases, project funding documents and announcements of support for open data innovation activities;
- Search for ‘hack days’, ‘open data boot camps’, ‘open data competitions’, ‘open data challenges’ ‘open data hackathons’ and check to identify if any of these were supported by government;
- Global open data innovation events and programmes such as the annual open data day [[http://wiki.opendataday.org/2015/City\\_Events](http://wiki.opendataday.org/2015/City_Events)];
- Research and innovation networks such as the Open Data Research Network [<http://www.opendataresearch.org/>] and the Open Data Innovations Network [<https://www.linkedin.com/grp/home?gid=4790214>];

- Regional open data communities, such as the ePSI Platform [<http://www.epsiplatform.eu/>], the Caribbean open data Institute [<http://caribbeanopeninstitute.org/content/open-data>] or the Africa Open Data [<https://www.facebook.com/groups/africaopendata/>], Open Data Francophone [<https://www.linkedin.com/grp/home?gid=8313551>], Open Government in the Arab States [<https://www.linkedin.com/grp/home?gid=4850889>] and Open MENA [<http://openmena.net/en/>] networks.

## C8) To what extent are city, regional and local governments running their own open data initiatives? [ODB.2013.C.CITY]

### Scoring Criteria

10	5	0
<p>All of the largest cities and regions in the country have open data initiatives. The majority of these have strong political backing and are well resourced. There is also open data presence in other smaller cities. Coordination between these initiatives and the national one is also expected</p>	<p>A number of different cities or regions across the country have open data initiatives. These may be predominantly run by civil society, or may have limited resources dedicated to them. Open data initiatives exist in cities and regions beyond the national or state capitals.</p>	<p>There are no city or regional open data initiatives.</p>

### Evidence and scoring thresholds:

**For a score of above 0** to be awarded: there should be evidence of at least some open data activity at the city or regional level, even if just incipient.

**For a score of above 3** to be awarded: there should be evidence of open data initiatives in at least the national or state capitals. These should be running their own open data initiatives with dedicated data catalogues.

**For a score of above 5** to be awarded: there should be evidence of a number of different cities or regions across the country with open data initiatives beyond the national or state capitals. These may be predominantly run by civil society, or may have limited resources dedicated to them.

**For a score above 8** to be awarded: there should be evidence of the presence of open data initiatives in most of the largest cities and regions in the country as well as in other smaller ones. The majority of these should have strong political backing and be well resourced with their own dedicated policies. Some level of coordination between these initiatives and the national one is also expected.

### Scoring Guidance

Open government data does not just involve central government. Regional, city and local government may all adopt open data initiatives. The criteria of an open data initiative are the same as in C1.

You can find a list of the largest cities by population for each country from the source information at [\[https://en.wikipedia.org/wiki/List\\_of\\_largest\\_cities\\_and\\_second\\_largest\\_cities\\_by\\_country\]](https://en.wikipedia.org/wiki/List_of_largest_cities_and_second_largest_cities_by_country).

For a score of 10 check the top-10 cities by population (10 largest States by population for federal systems), but also activity at some slower ones.

### Source Guidance

- Conversations with open data specialists in civil society organizations or individuals who are directing open data campaigns;
- Conversations with government officials working in open data offices or projects;
- Conversations with NGO officers with expertise in open-data and access-to-information issues, and investigative journalists;
- Searching for data portals and platforms run by cities or regions;
- Reports published by the media, academic and policy journals, and development and multilateral bodies (e.g., Open Knowledge Foundation, Sunlight Foundation, etc.)
- Regional open data communities, such as the ePSI Platform [\[http://www.epsiplatform.eu/\]](http://www.epsiplatform.eu/), the Caribbean open data Institute [\[http://caribbeanopeninstitute.org/content/open-data\]](http://caribbeanopeninstitute.org/content/open-data) or the Africa Open Data [\[https://www.facebook.com/groups/africaopendata/\]](https://www.facebook.com/groups/africaopendata/), Open Data Francophone [\[https://www.linkedin.com/grp/home?gid=8313551\]](https://www.linkedin.com/grp/home?gid=8313551), Open Government in the Arab States [\[https://www.linkedin.com/grp/home?gid=4850889\]](https://www.linkedin.com/grp/home?gid=4850889) and Open MENA [\[http://openmena.net/en/\]](http://openmena.net/en/) networks.

## C9) To what extent is training about open data available for individuals or businesses who want to increase their technical skills or develop businesses to use (open) data? [ODB.2013.C.TRAIN]

### Scoring Criteria

10	5	0
<p>There is widespread access to high quality training covering the full range of open data issues, such as data technology, data science and statistics, data visualisation and the business and legal aspects of open data. Both basic and advanced training are available from a range of different providers and in a range of cities and locations across the country. Much of the training provided is 'world class' and would equip those trained to operate in a global marketplace.</p>	<p>There is limited access to training on open data issues, such as data technology, data science and statistics, data visualisation and the business and legal aspects of open data. Whilst there may be some examples of training that would equip those trained to operate in a global marketplace, the majority of training may be of middling or low quality.</p>	<p>There is very limited training available, or those wishing to learn have to access training from outside the country.</p>

### Evidence and scoring thresholds:

**For a score of above 0** to be awarded: there should be evidence of some training available, but in a very limited way, mostly as informative sessions focused on general data and/or open data issues or mainly from outside the country (including online training).

**For a score of above 3** to be awarded: there should be evidence of limited access to training on general data issues, such as general data technology, statistics and data visualisations, but not specifically on the peculiarities of open data.

**For a score of above 5** to be awarded: there should be evidence of limited access to training on more specific data issues, such as data science, geographic information systems (GIS) and big data as well as open data issues, like the business and legal aspects of open data. Whilst there may be some examples of training that would equip those trained to operate in a global marketplace, the majority of training may be of middling or low quality.

**For a score above 8** to be awarded: there should be widespread access to high quality training covering the full range of data analytics and open data issues, as well as other thematic training

such as geodata or data journalism. Both basic and advanced training are available from a range of different providers and in a range of cities and locations across the country. Much of the training provided is ‘world class’ and would equip those trained to operate in a global marketplace.

### Scoring Guidance

Working with open data involves a wide range of knowledge and skills, including:

- **Web technologies** - covering issues such as web development, publishing data on the web, and building applications with open data for a range of platforms, including mobile;
- **Data science** - covering core statistical skills, and the ability to work with large online data sets from heterogeneous sources;
- **Data visualisation** - covering design and programming skills to create interactive and static data visualisations and infographics;
- **Legal aspects of open data** - addressing issues such as the licensing of open datasets, and legal issues in open source, open data and open knowledge;
- **Business aspects of open data** - addressing the creation of commercial enterprises on top of open data or open source - delivered through formal training or business mentoring and incubation programmes.

Training may be delivered through i), both full-time and part-time, or through professional development courses. Training may also be delivered through business incubator programmes, or short-term boot-camp training events. Training on general web and data technologies, data science and visualisations, and increasingly also on more specific open data issues, is usually available through university curriculums and programmes.

Where training events are one-off, or short-term donor funded interventions in one or two locations in the country, the maximum score available should be 5. Where training on the majority of the topics noted above is accessible to businesses and individuals in the country who may be interested in it, then a score of 10 should be given.

Where online training is available in languages accessible to businesses and individuals in the country, *and* there is evidence that this training is used by nationals of the country, then this may count towards your assessment of ‘limited access to training’ (5), but online only courses should not be taken to count towards ‘widespread access to training’ (10).

### Source Guidance

- Conversations with developers and technology businesses interested in developing their open data capacity to identify the training they access;
- Conversations with training providers about the courses offered and the range of courses available;



- Course directories and promotional material from further and higher education providers, including universities;
- Web searches for training providers on the particular topics noted above;

## OPEN DATA IMPACT QUESTIONS

I1 to I6 ask you to assess the degree to which there is any evidence that **open data release by the country government** has had impacts in a variety of different domains in the country.

The question Scoring Criteria invite you to look for credible claims made in academic and scientific publications, use cases, mainstream media and other accredited online sources which explicitly attribute certain impacts to open data released by the country government. **The highest scores are only available where there is peer-reviewed or audited evidence of impact.**

We anticipate that in the 3rd edition of the Barometer there will be still few countries that score the highest scores, and some countries may have very low scores on this question.

### A side note on ‘impact’

**Measuring impact is notoriously difficult.** Establishing a solid causal connection between open data and particular changes is clearly beyond the scope of a survey such as the Barometer. However, for the purpose of the Barometer, claims made in credible sources concerning possible impacts of open data are a useful proxy indicator of areas in which impact may be occurring, and to allow initial comparison between countries.

**Claims about impact go beyond simple descriptions of where open data has been used.** For example, a newspaper might report that open data has been used to create a bus timetables application, to visualise the budget of the country, or in a hackathon focussed on public services. It is only when they relate this to some other outcome, such as budget savings, or an increase in use of public transport, that an impact claim has been made.

## General Impact Scoring Guidance

With the exception of question I6, for each topic area **you should carefully check to ensure the impacts you cite can be reasonably attributed to open government data from the country in question and released by the national government.** For example, a crowdsourcing application created by an NGO to monitor public services may contribute to more efficient government, but if it does not use open data from government then, for the purposes of these questions, you should not include it in your analysis. By contrast, if a government publishes data on its spending, and there are cases of third-parties using this to highlight where government could use its resources more efficiently, this would count as a case where impact could be cited from open government data.

Where examples draw on open data from outside the country (e.g. an ICT based in the country is using open data from the World Bank, some other government, or some global dataset) or city level data you may report this in Justifications and Sources, but these examples should not count towards scores higher than 5, as higher scores are only available where the data being used comes from the national government’s open data releases.

The top score (10) should only be given where at least one credible peer reviewed article clearly attributes measurable impacts to open data *and* there are further cases of impact in other credible media sources and blog posts etc. The justification of the highest ‘widespread impacts’ scores should also include examples cited from **at least three different sectors** (e.g. health, education, transport).

## I1) To what extent has open data had a noticeable impact on increasing government efficiency and effectiveness? [ODB.2013.I.GOV]

### Scoring Criteria

10	5	0
Open data is widely cited to have made a significant contribution to government efficiency, with rigorous evidence to back these claims: for example a peer-reviewed study or government audit showing the contribution of open data to government efficiency.	Two or more cases in the media or credible online sources where open data is cited to have contributed to increased government efficiency	No evidence of impact. For the purpose of this indicator, if no relevant open government data is available the score should also be 0.

### Scoring Guidance

Open data could lead to improvements in government efficiency and effectiveness in a number of ways:

- By enabling government departments to better plan and target resources;
- By allowing outside actors to scrutinise government use of resources and highlight areas for savings;
- By enabling outside actors to build new services on top of open data which deliver more effective public services;
- By supporting collaboration between different government departments;
- By conducting data analytics to identify patterns and develop new data-driven approaches to more informed policy making and service delivery;

- By cross linking data between different public agencies to produce shared content, services and policies among administrations.

You may find other ways that open data impacts government efficiency and effectiveness in addition to these.

#### Source Guidance

- Conversations with open data experts, government officials, NGOs and media;
- Reports and case studies from government departments, international organisations and NGOs;
- Media reports from domestic or international media;
- Blog posts and forum discussions;
- Research and innovation networks such as the Open Data Research Network [<http://www.opendataresearch.org/>] and the Open Data Innovations Network [<https://www.linkedin.com/grp/home?gid=4790214>];
- Repositories of research journals, such as Google Scholar [<http://scholar.google.com>] or the Open Data Research Network Bibliography [<http://bibliography.opendataresearch.org/>];
- Repositories of open data impact case studies and use examples such as the Open Data Impact Map [<http://www.opendataenterprise.org/map/viz/index.html>] or those curated by the GovLab [<http://thegovlab.org/a-repository-of-open-data-repositories-a-repository-of-open-data-repositories-open-data-impact-case-studies-and-examples/>].

## I2) To what extent has open data had a noticeable impact on increasing transparency and accountability in the country? [ODB.2013.I.ACCOUNT]

### Scoring Criteria

10	5	0
Open data is widely cited to have made a significant contribution to increased transparency and accountability, with rigorous evidence to back these claims: for example, a peer-reviewed study showing a positive impact on transparency and accountability.	Two or more cases in the media or credible online sources where open data is cited to have contributed to greater transparency and accountability.	No evidence of impact. For the purpose of this indicator, if no relevant open government data is available the score should also be 0.

### Scoring Guidance

Open data could lead to improvements in government transparency and accountability in a number of ways:

- Supporting journalism and data journalism which uncovers wasteful spending, corruption or other wrongdoing by government departments or officials;
- Supporting the creation of applications which allow citizens to report on their experience of government services (for example, when a directory of schools or hospitals helps third-parties build a school or healthcare performance reporting application for citizens);
- Supporting scrutiny of government decision making;
- Supporting greater citizen engagement in policy making.

You may find other ways that open data impacts government transparency and accountability in addition to these.

This question is intended to evaluate whether open data is being used by civil society, individuals or non-government professionals to increase government transparency and accountability. Examples include applications to leverage open data around budgets, expenditure tracking, procurement, taxation and policy making, as well as accountability mechanisms such as financial disclosure, conflict-of-interest restrictions, audit systems and anti-corruption efforts.

Researchers should first check the websites of ministries such as health, education, transportation or agriculture for available open data. They should also check for national open

data portals that make data available to the community free of charge (e.g., Open Data Kenya, Healthdata.gov, Data.gov.uk). If there is available open data in a country, further investigation should be done to evaluate the extent that this data is used to improve government transparency and accountability.

#### Source Guidance

- Conversations with open data experts, government officials, NGOs and media;
- Reports and case studies from government departments, international organisations and NGOs;
- Media reports from domestic or international media;
- Blog posts and forum discussions.
- Research and innovation networks such as the Open Data Research Network [<http://www.opendataresearch.org/>] and the Open Data Innovations Network [<https://www.linkedin.com/grp/home?gid=4790214>];
- Repositories of research journals, such as Google Scholar [<http://scholar.google.com>] or the Open Data Research Network Bibliography [<http://bibliography.opendataresearch.org/>];
- Repositories of open data impact case studies and use examples such as the Open Data Impact Map [<http://www.opendataenterprise.org/map/viz/index.html>] or those curated by the GovLab [<http://thegovlab.org/a-repository-of-open-data-repositories-a-repository-of-open-data-repositories-open-data-impact-case-studies-and-examples/>].

### I3) To what extent has open data had a noticeable impact on environmental sustainability in the country? [ODB.2013.I.ENV]

#### Scoring Criteria

10	5	0
Open data is widely cited to have made a significant contribution to the environmental sustainability, with rigorous evidence to back these claims: for example a peer-reviewed study showing the impact of an open dataset on environmental sustainability.	Two or more cases in the media or credible online sources where open data is cited to have had an impact on environmental sustainability.	No evidence of impact. For the purpose of this indicator, if no relevant open government data is available the score should also be 0.

#### Scoring Guidance

Open data could lead to impacts on environment sustainability in a number of ways:

- Through enabling greater scrutiny of pollution impacts or environmental impacts of government projects or private enterprise;
- Through supporting greater attention to be paid to environmental factors in planning projects;
- Through encouraging government buildings to make more efficient use of energy;
- Through raising citizens awareness of their own environmental impacts;
- Through supporting campaigns on environmental issues.

You may find other ways that open data impacts environmental sustainability in addition to these.

You may include in your assessment open datasets which, whilst not directly published by government, are published as a result of a government mandate. For example, where a regulator may require that factories publish pollution statistics. However, you should check that the published data is **open data**, licensed for reuse in these cases.

This question is intended to evaluate whether open data is being used by civil society, individuals or non-government professionals to enhance environmental sustainability. Examples of areas where open data can be applied are alternative energy (solar, wind, hydro), energy-efficient construction, pollution (air, water, soil), water usage, waste management and sewage treatment, recycling programs and climate change.

Researchers should first check the websites of ministries such as health, education, transportation or agriculture for available open data. They should also check for national open data portals that make data available to the community free of charge (e.g., Open Data Kenya, Healthdata.gov, Data.gov.uk). If there is available open data in a country, further investigation should be done to evaluate the extent that this data is used to enhance environmental sustainability.

### Source Guidance

- Conversations with open data experts, government officials, NGOs and media;
- Reports and case studies from government departments, international organisations and NGOs;
- Media reports from domestic or international media;
- Blog posts and forum discussions.
- Research and innovation networks such as the Open Data Research Network [<http://www.opendataresearch.org/>] and the Open Data Innovations Network [<https://www.linkedin.com/grp/home?gid=4790214>];
- Repositories of research journals, such as Google Scholar [<http://scholar.google.com>] or the Open Data Research Network Bibliography [<http://bibliography.opendataresearch.org/>];
- Repositories of open data impact case studies and use examples such as the Open Data Impact Map [<http://www.opendataenterprise.org/map/viz/index.html>] or those curated by the GovLab [<http://thegovlab.org/a-repository-of-open-data-repositories-a-repository-of-open-data-repositories-open-data-impact-case-studies-and-examples/>].



## I4) To what extent has open data had a noticeable impact on increasing the inclusion of marginalised groups in policy making and accessing government services? [ODB.2013.I.INC]

### Scoring Criteria

10	5	0
Open data is widely cited to have made a significant contribution to the inclusion of marginalised groups, with rigorous evidence to back these claims: for example a peer-reviewed study showing the greater inclusion of at least one marginalised group.	Two or more cases in the media or credible online sources where open data is cited to have allowed marginalised groups to participate in either policy making, or accessing government services.	No evidence of impact. For the purpose of this indicator, if no relevant open government data is available the score should also be 0.

### Scoring Guidance

All societies have certain groups who are marginalised. This may be on grounds of age, gender, race, tribe, caste, class, disability, geographic location, and levels of poverty. Whilst these groups are not prohibited by licenses or technical mechanisms from accessing and using open data, they may not always be able to have effective access to open data. It has been argued however that open data can lead to more inclusive policy making and government services.

This may happen through the direct use of open data by marginalised groups, or through the work of intermediary organisations who support marginalised groups to access and use data, or who use data to campaign for the greater inclusion of marginalised groups in decision making or in receiving the benefits of public services.

### Source Guidance

- Conversations with open data experts, government officials, NGOs and media;
- Reports and case studies from government departments, international organisations and NGOs;
- Media reports from domestic or international media;
- Blog posts and forum discussions.
- Research and innovation networks such as the Open Data Research Network [<http://www.opendataresearch.org/>] and the Open Data Innovations Network [<https://www.linkedin.com/grp/home?gid=4790214>];

- Repositories of research journals, such as Google Scholar [<http://scholar.google.com>] or the Open Data Research Network Bibliography [<http://bibliography.opendataresearch.org/>];
- Repositories of open data impact case studies and use examples such as the Open Data Impact Map [<http://www.opendataenterprise.org/map/viz/index.html>] or those curated by the GovLab [<http://thegovlab.org/a-repository-of-open-data-repositories-a-repository-of-open-data-repositories-open-data-impact-case-studies-and-examples/>].

## I5) To what extent has open data had a noticeable positive impact on the economy? [ODB.2013.I.ECON]

### Scoring Criteria

10	5	0
Clear rigorous evidence of a contribution of open data to a range of forms of new economic activity and/or economic growth; for example, a peer-reviewed study showing a positive impact on economic growth across a number of sectors.	Multiple cases in the media or credible online sources where open data is cited to have contributed to economic growth in at least two different sectors.	No evidence of impact. For the purpose of this indicator, if no relevant open government data is available the score should also be 0.

### Scoring Guidance

Open data may impact on the economy in a number of ways. For example:

- Through supporting the creation of new businesses based on open data;
- Through supporting existing businesses to lower their costs or become more efficient (for example, using weather or transport data to better plan their operations)
- Through supporting better economic planning.

You can read about a range of open data business models at:

<http://radar.oreilly.com/2013/01/open-data-business-models-deloitte-insight.html>

There are many articles concerning the potential contribution of open data to economic growth, but to score 10, studies or credible sources cited should be about actual observed economic growth, rather than forecasts of potential economic impacts from open data.

### Source Guidance

- Conversations with businesses, open data experts, government officials, NGOs and media;
- Reports and case studies from government departments, international organisations and NGOs;
- Media reports from domestic or international media;
- Blog posts and forum discussions.

- Research and innovation networks such as the Open Data Research Network [<http://www.opendataresearch.org/>] and the Open Data Innovations Network [<https://www.linkedin.com/grp/home?gid=4790214>];
- Repositories of research journals, such as Google Scholar [<http://scholar.google.com>] or the Open Data Research Network Bibliography [<http://bibliography.opendataresearch.org/>];
- Repositories of open data impact case studies and use examples such as the Open Data Impact Map [<http://www.opendataenterprise.org/map/viz/index.html>] or those curated by the GovLab [<http://thegovlab.org/a-repository-of-open-data-repositories-a-repository-of-open-data-repositories-open-data-impact-case-studies-and-examples/>].

## 16) To what extent are entrepreneurs successfully using open data to build new businesses in the country? [ODB.2013.I.ENTR]

### Scoring Criteria

10	5	0
There are five or more examples of successful and thriving new businesses based around open data, employing staff and/or attracting significant external investment.	There are small scale examples of commercial apps, websites or other businesses built with open data. These remain niche or small scale businesses.	There is no evidence of entrepreneurial use of open data.

### Scoring Guidance

This question focuses specifically on entrepreneurial use of open government data. See I5 for details of business models around open data.

An entrepreneurial use of open data involves a company that:

- Earns revenue from its products and services, and
- Use open government data as a **key resource** for its business.

Be careful to distinguish business based on **open data** (explicitly citing certain open datasets as a key input into their work) from businesses based around open source software. Open source software is distinct from open data, and open source businesses do not count towards scores for this question.

### Source Guidance

- Conversations with businesses, open data experts, government officials, NGOs and media;
- Reports and case studies from government departments, international organisations and NGOs;
- Media reports from domestic or international media;
- Blog posts and forum discussions;
- Research and innovation networks such as the Open Data Research Network [<http://www.opendataresearch.org/>] and the Open Data Innovations Network [<https://www.linkedin.com/grp/home?gid=4790214>];
- Repositories of research journals, such as Google Scholar [<http://scholar.google.com>] or the Open Data Research Network Bibliography [<http://bibliography.opendataresearch.org/>];

- Repositories of open data impact case studies and use examples such as the Open Data Impact Map [<http://www.opendataenterprise.org/map/viz/index.html>] or those curated by the GovLab [<http://thegovlab.org/a-repository-of-open-data-repositories-a-repository-of-open-data-repositories-open-data-impact-case-studies-and-examples/>];
- The repository of open data driven companies in emerging markets [<https://finances.worldbank.org/dataset/Pipeline-research-open-data-driven-companies-in-em/u4wq-7b6m>].

## OPEN DATASETS QUESTIONS

To assess how far open data policies are being implemented the following questions ask you about whether government makes available data on a range of different topics in this country, and where the best sources for that data may be.

### Search tips: locating datasets online

1. **If the government has a central open data portal search there first.** If you are not sure if the country has an open data portal search the web for relevant mentions. You might also find data portals from specific departments, particularly national statistics agencies.

You can also find two extensive lists of known Open Government Data portals at <http://www.opengeocode.org/opendata/> and <http://dataportals.org/> **Note:** neither of these sources are authoritative, and may contain outdated data, or may miss out on data, particularly information not accessible in English. **You should confirm for yourself any information on datasets gathered or missing from these sites.**

2. **Check the websites of any government departments or agencies who may be responsible for this dataset** - use their internal search features to look for the relevant data or to search for 'open data' to see if they contain a listing of available datasets.

*Consider using the 'site:' and 'filetype:' operators in Google search as explained below to probe the websites if their in-built search is not very effective.*

3. **Search the web for mentions of this data or any other pointers to a possible dataset.** In addition, to identify whether a dataset is available, and available as open data, you might:
  - Enquire with the relevant government department directly;
  - Ask on the mailing lists of the local open data community if one exists;
  - Consult academics who work with this kind of data to identify whether they have found sources for it.

### Using 'site:' and 'filetype:' operators.

Google search has two advanced search options that can help you locate data.

**1. The 'site:' operator restricts search to files indexed by Google on the site(s) you specify.**

For example, having located the website of the Geospatial agency in Indonesia I can run the search:

[site:bakosurtanal.go.id "open data"](#)

to search for mentions of the phrase "open data" anywhere on the website. I could also widen the search to check on any government website using 'go.id' as the site. For example:

[site:go.id "open data"](#)

**2. Using the 'filetype:' operator I can look for files of a particular kind. For example:**

[site:bps.go.id filetype:csv](#)

searches the Statistics Indonesia website for any CSV files that may contain relevant statistical data. To explore whether a particular dataset might be available I could guess at something I would expect to find within it and add this to my query.



## DATA CATEGORIES

Below you will find a list of dataset categories to be assessed, along with an operational definition of those categories. Examples are given of data that might fit into this category. The examples are not necessarily 'open data', but simply indicate the kinds of data that would meet the requirements of the question.

### D1. Map Data (full coverage of the country)

**A detailed digital map of the country** provided by a national mapping agency **with key features such as official administrative borders, roads and other important infrastructure**. Please look for **maps of at least a scale of 1:250,000 or better (1cm = 2.5km)**.

#### Examples:

The United Kingdom mapping agency, the Ordnance Survey, provides open data mapping products for download:

<http://www.ordnancesurvey.co.uk/oswebsite/opendata/discover.html>

The United States Census.gov site provides detailed mapping products for download:

<http://www.census.gov/geo/maps-data/>

The Mozambique mapping agency at

[http://www.cenacarta.com/modules.php?name=Downloads&d\\_op=viewdownload&cid=5](http://www.cenacarta.com/modules.php?name=Downloads&d_op=viewdownload&cid=5)

provide shape files for download.

### D2. Land Ownership Data

**A dataset that provides national level information on land ownership**. This will usually be held by a land registration agency, and usually relies on the existence of a national land registration database.

#### Examples:

The New Zealand Land Registry makes data on 'land parcels' available as open data as listed at <http://data.govt.nz/dataset/show/1333>

The UK Land Registry publishes monthly residential sales 'Price Paid' data on land transactions at

<http://www.landregistry.gov.uk/market-trend-data/public-data/price-paid-data>

This [news story](#) reveals that Delhi has implemented a GIS system for land ownership, but also implies that a national land ownership database does not currently exist.

### D3. Question intentionally left blank

*There is no D3 question in the survey.*

### D4. National Statistics

**Key national statistics including demographic and/or economic indicators** (GDP, unemployment, population, etc), often provided by a National Statistics Agency. **Aggregated data is considered acceptable for this category** for the time being (e.g. GDP for whole country at a quarterly level, or population at an annual level).

Examples:

The [Moldovan Open Data Portal](#) contains datasets on population distribution drawn from the Census.

[Detailed bulk data tables](#) from the UKs 2011 Census are available online.

[Statistics Canada](#) provide in depth census tables and other national statistics for download.

### D5. Detailed government budget

**National government budget at a high level** (e.g. spending by sector, department etc). Budgets are government plans for expenditure, (not details of actual expenditure in the past which is covered in the spend category).

Research tips:

The Open Budget Survey (not universal coverage) regularly assesses whether budget data is available from a government [<http://survey.internationalbudget.org/>].

The [OpenSpending.org](#) platform have brought together many budget datasets, and may help you locate whether the government is publishing open budget data or not. In some cases data is being made available on open spending by third-parties converting from printed or PDF budget records. In these cases you can judge that data is available, but should decide whether or not this is sustainable, given the effort required on the part of civil society.

The World Bank have published a detailed report and 'FMIS World Map' dataset on Integrated Financial Management Information Systems in government. See [\[http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPUBLICSECTORANDGOVERNANCE/0,,contentMDK:23118772~pagePK:148956~piPK:216618~theSitePK:286305,00.html\]](http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPUBLICSECTORANDGOVERNANCE/0,,contentMDK:23118772~pagePK:148956~piPK:216618~theSitePK:286305,00.html) and an Open Budgets dedicated portal [\[http://wbi.worldbank.org/boost/\]](http://wbi.worldbank.org/boost/)

#### Examples:

South Africa publishes detailed information on the countries budget online at <http://www.treasury.gov.za/documents/national%20budget/default.aspx> - a search on the site for xls files (using google> 'site:treasury.gov.za filetype:xls budget') suggests that datasets for elements of the budget may also be available, although further investigation is needed to confirm.

The UK Data Portal lists the PESA budget dataset at <http://data.gov.uk/dataset/hmt-pesa09-chapter1>

## D6. Detailed data on government spend

**Records of actual (past) national government spending at a detailed transactional level; at the level of month to month government expenditure on specific items** (usually this means individual records of spending amounts under \$1m or even under \$100k). Note: A database of contracts awarded or similar is not sufficient for this category, which refers to detailed ongoing data on actual expenditure.

#### Examples:

The Uruguay Office of Planning and Budget publish a visualisation of public spending, along with underlying datasets: <http://agev.opp.gub.uy/advni/>

The UK has mandated detailed publication of all government spend over £25,000 and over 500 spend datasets can be found at data.gov.uk: <http://data.gov.uk/data/search?tags=spend-transactions>

## D7. Company register

**A list of registered (limited liability) companies in the country including name, unique identifier and additional information such as address and registered activities.** The data in this category does not need to include detailed financial data such as balance sheet, etc.

#### Research tips:

The Wikipedia list of company registrars by country may prove useful in research:  
[http://en.wikipedia.org/wiki/List\\_of\\_company\\_registers](http://en.wikipedia.org/wiki/List_of_company_registers)

[OpenCorporates.com](http://opencorporates.com) have scraped data from many different corporate registries and maintains an Open Company Data Index [<http://registries.opencorporates.com>]. Each record points you to details of the government held data, linking back to it's original source of information where you could double check such details. If the data only appears to be available in bulk form on Open Corporates, and is not available in bulk from the government, then do not award the dataset a mark for sustainability.

Examples:

[UK Companies House](http://companieshouse.gov.uk) list a 'free public data product' on their company information pages, which consists of a download of basic company information.

## D8. Legislation

**The constitution and laws of a country.** Please detail whether the data source covers all the laws of the country, or only a sub-set thereof.

Research tips:

Legislative open data will often be provided using eXtensible Markup Language (XML), so searching for '[country] legislation XML' may help locate open legislation data.

Examples:

The UK [Legislation.gov.uk](http://legislation.gov.uk) site provides laws as structured data as described in this [Legislation as open data](#) presentation. The site provides stable URLs for laws.

XML data files of recent US Laws are available through the Library of Congress Thomas site, as detailed at <http://xml.house.gov/>.

## D9. Public transport timetables

**Details of when and where public transport services, such as buses and rail services, are expected to run.** Please provide details for both bus and rail services if applicable. **If no national data is available, please check and provide details related to the capital city** (clearly stating that in your justification as well).

Research tips:

Sometimes transport information is provided through an ‘API’ (Application Programming Interface) rather than as bulk downloads. The GTFS data exchange for example provides details of many local areas sharing public transport data <http://www.gtfs-data-exchange.com/agencies/bylocation>

Examples:

[The Traveline National Dataset of public transport timetables](#) provides UK public transport timetables.

Labs Ruter provide an [API onto real-time transport information for Norway](#), released under an open license.

## D10. International trade data

**Details of the import and export of specific commodities and/or balance of trade data against other countries.**

Examples:

Data.gov.uk lists [a range of trade statistics available](#) under the Open Government License.

The Central Statistics Organisation of Afghanistan publish [a list of exports by commodity and country](#) in a Excel spreadsheet on their website, along with other trade statistics.

## D11. Health sector performance

**Statistics generated from administrative data that could be used to indicate performance of specific services, or the healthcare system as a whole.** Health performance data might include:

- Levels of vaccination;
- Levels of access to health care;
- Health care outcomes for particular groups;
- Patient satisfaction with health services.

**Simple lists of health facilities do not qualify as health performance data.**

Research tips:

[The Global Health Data Exchange](#) maintain country profiles and a repository of datasets for countries across the world. The [advanced search](#) can be used to find 'Administrative Record' data types'. This can help you locate national sources of health data, and explore whether these are open datasets. Not all the sources listed on the Exchange are from national governments, so check carefully when exploring available data.

National ministry of health websites are a useful starting point for a health data search in countries that do not have a national open data portal.

#### Examples:

The [UK Department of Health](#) publish extensive statistics on the performance of the health service.

The [Kenya Open Data Portal includes](#) statistics on levels of immunisation in the country, although the data is not updated.

## D12. Primary and secondary education performance data

**Statistics generated from administrative data that could be used to indicate performance of specific services, or the education system as a whole.** Performance data might include:

- Test scores for pupils in national examinations;
- School attendance rates;
- Teacher attendance rates.

**Simple lists of schools do not qualify as education performance data.**

#### Examples:

The [Kenya Open Data Portal](#) includes a range of education datasets, including attendance and test score statistics.

Brazil [publishes a microdata census of schools annually](#).

[The New Zealand Open Data Portal](#) lists a range of education and learning outcomes datasets, including primary and secondary school achievements.

## D13. Crime statistics

**Annual returns on levels of crime and/or detailed crime reports.** Crime statistics can be provided at a variety of levels of granularity, from annual returns on levels of crime, to detailed

real-time crime-by-crime reports published online and geolocated, allowing the creation of crime maps.

#### Research tips:

Look for data from a national criminal records agency, or policing department.

#### Examples:

The Indian [National Crime Records Bureau](#) publish annual crime and prison statistics.

The [UK Police Force publishes street level](#) crime data on a monthly basis, sharing the location of individual crime reports.

### D14. Environmental data

**Data on one or more of: carbon emissions, emission of pollutants (e.g. carbon monoxides, nitrogen oxides, particulate matter etc.) and deforestation.** Please provide links to sources for each if available.

#### Examples:

The [UK Department for Food and Rural Affairs](#) maintain an archive of air quality information.

The [Mexican 'Your Government Maps' service](#) provides maps on a variety of environmental indicators.

### D15. National election results

**Results by constituency / district for the most of all national electoral contests over the last ten years.** If only partial data is available please provide details.

#### Examples:

The [Swedish Election Authority](#) provide detailed downloads of election information, returns and results.

The [UK Electoral Commission](#) publish a spreadsheet of recent election results.

### D16. Public contracting data

**Details of the contracts issued by the national government. When answering this question please look for sources that provide contract award data and not only requests for bids** (i.e. details of the fact a contract has been put in place) If only solicitation/tenders are available, please note this in your description of the available data.

#### Research tips:

Governments enter into many contracts for the provision of goods and services. Through websites publishing tenders, government contract finder websites, procurement portals and contract websites they may make information about the tender (request for bids) and award (details of who received the contract and it's value) online.

Look for services that aggregate data from across government - not just single departmental websites. However, if no such service is available, check a selection of the biggest government departments and note if they publish their contract data in any form.

When looking for **open data** from a contracts portal, look for options to 'export' searches or feeds. If a national data portal exists, check there for national government contract data.

#### Examples:

[The United Kingdom Contracts Finder website](#) provides information on tenders and awarded contracts. It provides data feeds and on the data.gov.uk website there are details of a regularly updated data dump for this site.

## DATASET QUESTIONS

**For each dataset in the survey you will need to complete a ten points checklist.** You should also cross-reference answers after against:

- **The 2014 Open Data Barometer** - an interactive graph of country assessments is available at <http://opendatabarometer.org/assets/odb-datasets/2014.html>
- **The Global Open Data Index** - though note that this may have both false positives and false negatives <http://index.okfn.org/place/>

The dataset questions are as follow:

### a) Does the data exist?



**If the data is not collected in this country at all, or is not aggregated at a federal level in any form, then answer NO to this question.** You will be asked to provide an explanation of why the data does not exist and need to ensure that you have provided a detailed explanation of the search you undertook and sources consulted. This should include primary research, and not just finding a dataset was judged non-existent in existing third-party listings.

**If the data is collected at the level of government you are answering for in any form, then you should answer YES** (e.g. paper or digital). You will be asked to provide details of the agencies responsible for this data.

Research tips:

**If the data is not held by government at all, you will need also to explain briefly why the government does not hold this data** when possible (e.g. lack of capacity to collect the data, or the data in question being managed at a local rather than national level) **and whether the data may still be available from any other non-government third party.**

**If the data is held by government, you will need to provide a brief description of the data being provided** (i.e the different elements in the dataset and any known limitations), the **name of the agency or agencies that are responsible for the data** and the **link to where the data can be found.**

Examples:

- If you are answering for '**census data**' and a census is carried out, with records kept on paper, then answer **YES**.
- If you are answering for '**census data**' and no data is collected, or the data is only collected in one or two regions, answer **NO**.

## b) Is it available online from government in any form?

**Can you find the data online in any digital format?** This might include tables on web pages, PDF files, or scanned copies of paper documents, as well as other formats like Excel.

**You should only answer YES if a substantial proportion of the data itself is available.** Press releases or short summaries selecting just a few of the possible data points do not count. Under each particular dataset guidance is provided on what might constitute adequate online data.

Research tips:

**Ensure the source you find is the responsibility of government.** A government domain name, or clear statement that the website is sponsored by government may provide evidence of

this. If you find a dataset available on a non-government website then further searching to find if a government source is also available may be required. **If no government source is available, answer NO to this question, but** continue to the following questions based on any non-government source you may have located.

Examples:

- If looking for **mapping data** and the government has published **image files containing maps with boundaries and locations of key infrastructure** (e.g. roads, rail) then answer **YES**.
- If looking at **census data** and the only data available is in **PDF reports with tables and graphs, but without in-depth statistical tables covering the majority of main census indicators**, then answer **NO**.

You will also be asked to provide details of whether the data, of equivalent data, is available **as open data** from a third party. For example, in the case of mapping data, you could check for good [Open Street Map](#) coverage for the country.

### c) Is the dataset provided in machine readable and reusable formats?

**Can you find a copy of the data which can be opened in appropriate data manipulation and analysis software?** If there are multiple datasets that you could answer with respect to (for example, when a government publishes different data products based upon a census or land registry, or when there are a number of different government service directories maintained) **you should answer with respect to the best example you can locate in terms of availability of open data.**

For example, if the government holds directories on hospitals, community health centres and doctors, but only the list of hospitals is in a machine readable form, you should note this in the justification and continue to answer the questions with respect to this hospitals' dataset.

Research tips:

You may need to search the site where you find the data for a machine readable form. For example, if on a data portal, go back to the root of the portal and search for all mentions of this dataset to find different versions. If on a government website, use the 'site:' operator to search the site for 'open data', 'datasets' or files of csv, xls or other types.

Examples:

- If the data is only available as **scanned copies of printed out documents in PDF or in image formats** answer **NO**. In these cases the only way to get data that can be sorted

or manipulated would be to retype it into a spreadsheet. This shows the data was not originally machine readable.

- If looking at **statistical data**, and there is a version of the dataset you can load into **spreadsheet** software and sort and filter, answer **YES**.
- If looking at **statistical data** and the only data available is in **PDF files, or on (HTML) web pages** where copying and pasting the data into a spreadsheet does not let you easily sort and filter it, answer **NO**.
- If looking for **mapping data** and there are **shapefiles (SHP), Points of Interest (KML)** or other geographic data formats provided, answer **YES**.
- If looking for **mapping data** and the only data available is in the form of **images**, even if in 'slippy maps' online, answer **NO**.
- If there is an **API (Application Programming Interface) for the data**, or **bulk downloads provided in XML, CSV or JSON formats**, answer **YES**.

#### d) Is the machine readable data available in bulk?

Bulk access to data files allows analysts and developers to more flexibly build upon the data and to integrate it into other products, services and activities. **Look for links to bulk downloads. These might be organised by month or year for time-series data, or broken down into sub-files for very large datasets.**

Research tips:

Sometimes it is only possible to access small extracts of a dataset using an online interface, as when, for example, you can only get a spreadsheet download after you build a query onto the dataset, or when data is scattered across tables or a very large number of download files on different web pages. However, **to qualify as bulk data it should be possible for a developer to easily collect a list of download file URLs and then grab this data and feed it into an application.**

Examples:

- If you are looking for *crime data* and the relevant agency publishes **a list of monthly crime statistics files for download on one single web page** answer **YES**.
- If looking for *transport timetable* data and the only way to get at a timetable is to **fill in an online form** for each transport route you are interested in, answer **NO**.

#### e) Is the dataset available free of charge?

**Sometimes datasets are only available for a charge - either a one-off payment or a subscription fee.** For example, when maps are licensed by the mapping agency and only

provided for a fee, or when a country's legislation is compiled and provided for sale through a private company.

#### Research tips:

If a bulk dataset is available, answer with respect to this. If the data is available online, but not in bulk, answer with respect to the online accessible version of the data. If a dataset is only available for a fee from a private company then please include notes on this in the comments.

#### Examples:

- If looking for **school performance data** and the data is **published for free in bulk on a government agency website**, answer **YES**.
- If looking for **legislation** and **a company sells bulk access to a database of laws on CD for a price**, and the data is not otherwise freely available, answer **NO**.

### f) Is the data openly licensed?

**Unless the license that a dataset is provided under is explicitly stated, users of the data cannot be sure whether they have permission to re-use it**, and under what conditions. For example, government datasets may be subject to copyright or intellectual property restrictions that make it illegal for a campaign group or company to build on top of the data or redistribute it.

#### Research tips:

If a bulk dataset is available, answer with respect to this. If the data is available online, but not in bulk, answer with respect to the online accessible version of the data. Look for a clear license statement setting out the permission that a user has. This will only count as an open license if:

- It clearly states that **anyone has permission to re-use it**;
- It **does not restrict what the data can be re-used for** (for example, through non-commercial restrictions) more than attribution and share-alike;

Refer to the [Open Definition](http://opendefinition.org/licenses/) and the list of conformant licenses [<http://opendefinition.org/licenses/>] for a detailed overview of what counts as an open license.

#### Examples:

- If there is **no clear statement concerning how the data can be re-used**, either next to the download, in the terms and conditions of the website, or on other relevant pages where the dataset can be found, answer **NO**.

- If there is a **copyright or license statement that states the dataset can only be used for non-commercial use, or that it is only for ‘personal’ or ‘research’ use** then answer **NO**.
- If there is a **clear open data license statement associated with the dataset** such as [Creative Commons Zero](#) (or CC-BY and CC-BY-SA), [Open Database License](#) (or ODC-BY), or [Open Government License](#) answer **YES**.
- If a **statement of general permissions is associated with the dataset, and the intent of the statement is clearly and unambiguously intended to permit re-use** of the data for all purposes, including commercial use, answer **YES**.
- If **there is statement of permissions, but it is ambiguous**, and no other license is given, answer **NO**.

### g) Is the dataset up to date?

Many datasets are only useful if they are recent and kept updated. **Consider any evidence on what was the last update date and how often it is being updated and how often you would expect data of this form to be updated.**

Research tips:

This assessment should focus on **the update of the dataset itself and not any meta-data records** associated with the data.

Examples:

- If looking at a **list of public services**, and there is a bulk data file for download, but it appears to have been **generated once-off, and not updated since**, answer **NO**.
- If looking at **census data**, and there are files from the **last census** conducted in the country, answer **YES**.
- If looking at **budget data**, and there are data files from the **most recent budget**, answer **YES**.

### h) Is the publication of this dataset sustainable?

**Look for evidence that this was a once-off publication, or for evidence that the dataset is being kept regularly updated** (for example using past track-record of publication).

Research tips:

Sometimes open data appears to involve just one-off publication of datasets, without any process in place to keep the data updated. Offer your best judgement as to whether it is likely

that this open data will be kept updated for at least a further 12 months (or longer if the update cycle for this data is longer than 12-months as in the case, for example, of election data, where you would judge on the basis of whether it is likely that open data will be published on results at the next election).

In some cases the evidence available to make a judgement will be quite limited. In these cases, **where you cannot locate any evidence that the data will be updated regularly**, you can answer **NO**.

Examples:

- If this is a type of data set that needs regular updates, but **only one version has been produced**, answer **NO**.
- If **all the datasets on a data portal were added on one data over 6 months ago, and there has been no activity since**, answer **NO**.
- If there appear to be **regular updates** to the dataset, for example, monthly versions posted on the web page, answer **YES**.
- If there is a **clear commitment from the responsible agency** to keep their open data updated, answer **YES**.

### i) Was it easy to find information on this dataset?

This needs to be answered thinking particularly about whether it was easy to find information on this dataset to answer questions a) to f) In addition to answering these supplementary questions **you should also add a note outlining the steps you took to search for this category of data**.

Research tips:

Answer **YES** if you think a **regular Internet user with a degree level education would be able to locate and find out about this dataset** if they were looking for it. Answer **NO** if you struggled to locate the dataset, or if you think a **degree educated Internet user in your country would face considerable struggles in trying to find and find out about this dataset**.

Examples:

- If once you located the dataset **all the information you needed about it's status as open data was just a few clicks away**, answer **YES**.
- If you **had to repeatedly go back to search engines to find out about dataset licenses etc**, or if you had to draw upon privileged information or contacts to find out about the dataset, answer **NO**.

## j) Are data URIs provided for key elements in the dataset?

**Datasets usually contain many different entities.** For example, a dataset on education might include record on about schools, regions and types of education programme. Advanced publication of these as open data, following [linked data principles](#) would involve creating a **URI/web address for each of these entities**, such that when a human or machine looked up that address they would be able to find data about that entity. The advantage of this is that when different datasets use the same URIs to identify a thing then connections can be more easily made between those datasets.

### Research tips:

URI stands for Uniform Resource Identifier. For the purpose of this question you can treat these the same as URLs (Uniform Resource *Locators*), or, as they are more commonly known, web addresses (e.g. <http://example.org/barometer>)

**First**, check to see whether a **‘Linked Data’ version of the dataset is advertised**. This may be provided in the ‘RDF (Resource Descriptor Format)’ file format, or as a file with the **.ttl, .rdf, or .n3 file extension**. To answer yes to this question these formats are *not* required, but **if you find these formats** then it is likely that the answer is **YES**.

Another clue that linked data version of this dataset are available is when a **‘SPARQL endpoint’ is mentioned**. SPARQL is a query language for linked data. These endpoints let people run their own queries against the data.

**Secondly**, if you don’t find these formats apparent, look to see if **there are unique web pages provided for all of the key entities in the dataset** *provided by* the organisation or agency responsible for the dataset (Note: third party publication of URIs does not count, unless under agreement with the government organisation/agency). For example, the UK Legislation website at <http://legislation.gov.uk> provides unique URIs for every bit of UK legislation, such as <http://www.legislation.gov.uk/ukpga/2012/9/contents/enacted>

If you find URIs about key entities from the dataset, then you can **check to see if a machine-readable data representation of the entity is available**. Look for **links to RDF, XML, JSON or other data formats** on the page. You may also ‘view source’ of the page, and check for `<link rel="alternate" type="X" ...>` links in the header which have **formats such as application/xml, application/json, text/csv** or similar.

### Examples:

- If queries for the dataset name and RDF, Linked Data and SPARQL provides no clear pointers to available data, answer **NO**.



- If you find **URIs for key entities from the dataset**, maintained by the agency responsible for the dataset, **and any machine-readable kind of data is available from these pages** (CSV, JSON, XML etc.) answer **YES**.
- If you find **a series of web pages for each entry in the dataset** (e.g. each school; each road) **and this has a link to a .rdf version**, which launches a file download, answer **YES**.
- If you search and find **a SPARQL endpoint which claims to contain the dataset**, and this appears to be **maintained by government**, or under agreement with government, answer **YES**.
- If you search and find **an RDF version of the dataset, or a SPARQL endpoint** which claims to contain the dataset, but this appears to have been **built by a third party unrelated with government**, answer **NO** (this is because in these cases the URIs given to entities are not likely to be being maintained by government).

#### Extra resources on Linked Data:

- Wikipedia page on Linked Data: [http://en.wikipedia.org/wiki/Linked\\_data](http://en.wikipedia.org/wiki/Linked_data)
- APARSEN - [Interoperability objectives and approaches](#) - 2.3.2 Coolness vs Persistence: Persistent Identifiers and Linked Data