A proposal of Global Open Data Index for the game of Go

Leonardo Alberto Dal Zovo Angela Corbari

Abstract. The use and reuse of open government data is constantly increasing not only for transparency and accountability but also for increasing inclusion, participation, engagement and empowerment of citizens. Starting from the work and experience of the Global Open Data Index by Open Knowledge, the proposal is to establish a Go Global Open Data Index to collect and present information on the current state of Open Data release of Go countries around the world in order to encourage National Go Organizations, Go clubs and Go players to work towards improving the quality and increasing the quantity of Open Data.

1 Introduction

The philosophy behind Open Data[1] is well established and its goals are similar to those of the other open movements, as Open Source for software. The concept of Open Data is spreading more and more nowadays, gaining popularity with the launch of Open Data government initiatives such as data.gov and data.gov.uk.

Open Knowledge[2], formerly known as Open Knowledge Foundation (OKF), is a nonprofit organization that promotes open knowledge including open content and Open Data. The organization published the Open Knowledge Definition and runs several projects, such as CKAN, the data portal software used by many governments for their Open Data projects, and Open Data Index, a website that shows measures and benchmarks on the openness of data around the world.

The proposed project aims to apply the Open Data Index project in the world of the game of Go.

[1] http://opendefinition.org/

[2] https://okfn.org/

2 Global Open Data Index by Open Knowledge

The Global Open Data Index[3] is a project that measures and benchmarks the openness of data around the world, and then presents this information in a way that is easy to understand and use. Each year the Open Data community and Open Knowledge produces an annual ranking of countries, peer reviewed by the Open Knowledge network of local Open Data experts.

Launched in 2013, the first edition of the Index reviewed the state of open government data in 70 countries, with over 1,300 submissions and over 60 volunteer country editors analysing over 700 government data sets. In 2014 it expanded to 97 countries with a focus on countries of the Global South.

Each year, governments are making more data available in an open format. The Global Open Data Index tracks whether this data is actually released in a way that is accessible to citizens, media and civil society and is unique in crowd-sourcing its survey of Open Data releases around the world.

The Global Open Data Index is not only a benchmarking tool, it also plays a powerful role in building the open government data community around the world. Governments and Open Data practitioners can review the Index results to see how accessible the Open Data they publish actually appears to citizens, and where improvements are necessary to make Open Data really open and useful. This increases its usefulness as an advocacy tool and broadens its impact.

2.1 Key questions

Global Open Data Index focus on ten key data sets[4] including those essential for transparency and accountability (such as election results and government

[3] http://index.okfn.org/

[4] http://index.okfn.org/dataset/

spending data), and those vital for providing critical services to citizens (such as maps and transport timetables):

- Transport timetables
- Government budget
- Government spending
- Election results
- Company register
- National map
- National statistics
- Legislation
- Postcodes/zipcodes
- Emissions of pollutants

2.2 Data collection[5]

Each dataset in each place is evaluated using nine questions that examine the technical and the legal openness of the dataset. In order to balance between the two aspects, each question is weighted differently and worth a different score. Together, the six technical questions are worth 50 points, the three legal questions are also worth 50 points.

The following questions examine technical openness:

- Does the data exist?
- Is the data in digital form?
- Is the data available online?
- Is the data machine-readable?
- Is it available in bulk?
- Is the data provided on a timely and up to date basis?

The following questions examine the legal status of openness:

- Is the data publicly available?
- Is the data available for free?
- Is the data openly licensed?

[5]Brief summary of the "Scoring" section available at `http://index.okfn.org/methodology/`

2.3 Assessment and quality review process[6]

The assessment takes place in two steps. The first step is collecting the evaluation of datasets through volunteer contributors, and the second step is verifying the results through volunteer expert reviewers. The following steps are taken each time a dataset is submitted:

1. Contributors, which can be any person, submit information about the availability of one of the key datasets in their Place. At this stage, the results are not published online straight away. Instead it is held back for review (see below).
2. Next, the submission is sent to a reviewer who has been appointed by Open Knowledge based on the person's experience and insight into Open Data in their Place or region. He or she will verify the results and make sure they are accurate. In cases where it was not possible to find a reviewer for a Place, we used an Open Knowledge staff member to review the results.
3. Once the review process is done, the entries are then scrutinized by a panel of expert reviewers, who are volunteers that carry particular Open Data expertise across the key dataset fields. This panel made a review thematically across several Places (vertical review in the table as compared to the prior reviewer's horizontal review).
4. Throughout the process, and right up until the day before the finalized Global Open Data Index was released under embargo to journalists, revisions could be submitted at any time by anyone who suggested having more updated or correct information about a dataset. Such contributions were then reviewed in the same manner as outlined above before being posted on the Index. This iterative process was put in place to allow as many people as possible to have a say.

3 The proposal for the Go Global Open Data Index

Starting from the experience of Global Open Data Index by Open Knowledge, the proposal aims to model the Go Global Open Data Index to collect and present information on the current state of Open Data release of Go countries around the world. It will not be a representation of the official Open Data offering in

[6]Brief summary of the "Assessment and quality review process" section available at `http://index.okfn.org/methodology/`

each country, but an independent assessment from a Go players' perspective.

The proposal aims to review the state of Open Data in all the International Go Federation's General Members (74 as of 2012)[7]. After the first year, based on the results of the first edition, it will be decided every when and how to repeat the census.

The following proposal for the Go Global Open Data Index will largely follow the instructions and directions provided by Open Knowledge for the Global Open Data Index. It will also make use of the same software platform licensed by Open Knowledge under the AGPL, in particular:

- *Open Data Census*[8], a webapp for doing Open Data Censuses including submission workflow, presentation of results and some visualization
- *Open Data Index*[9], it displays a snapshot of data collected in an Open Data Census

3.1 Key questions

The first Go Global Open Data Index will focus on eight key data sets, chosen based on the authors' experience and knowledge. The first five will be borrowed and partially adapted from the Global Open Data Index by Open Knowledge[10] and are strictly related to transparency and accountability, the last three, instead, will be introduced to model the special case of the game of Go (but they could be used also for other games):

1. *National Go Organization budget*: at a high level (e.g. spending by sector, department etc). This category is about budgets which are plans for expenditure (not actual expenditure in the past)
2. *National Go Organization spending*: records of actual (past) spending at a detailed transactional level. This data category refers to detailed ongoing data on actual expenditure
3. *Election results*: results for all major electoral contests
4. *National Statistics*: key national statistics such as players' population and their geographical distribution. Aggregate data is also considered acceptable in this data category
5. *Legislation (laws and statutes)*: this data category requires all regulations and statutes to be available online, although it is not a requirement that information on regulatory behaviour e.g. voting records is available
6. *Go clubs register*: list of recognised go clubs including name, unique identifier and additional information such as address and main contact points. The submissions in this data category does not need to include detailed financial data such as balance sheet etc
7. *Go players register*: list of recognised players including name[11], unique identifier and rank. The submissions in this data category does not need to include detailed personal data protected by privacy regulations.
8. *Tournament calendar*: list of recognised tournament including name, dates, place, unique identifier and additional information such as main contact point, website. Results, pictures or web links should be provided for past events

3.2 Data collection

Each dataset on each place is evaluated using the same nine questions of the Global Open Data Index by Open Knowledge[12].

A different balance of the two aspects will be used to determine the total score for a place, prioritising the availability of the data over the openness: as Open Data is in an introductory phase in the domain of the game of Go, it seems natural to give more importance at the beginning to the technical side of the data over the legal aspects.

Together, the six technical questions are worth 80 points, the three legal questions are instead worth 45 points. Each question will be weighted differently and worth a different score as follows:

- The following questions examine technical openness:
  - *Does the data exist?* 10 points
    Does the data exist at all? The data can be in any form (paper or digital, offline or online etc). If it is not, then all the other questions are not answered
  - *Is the data in digital form?* 10 points
    Can the data can be accessed from a computer or is it in the old paper form? If you

[7] `http://intergofed.org/about-the-igf.html`

[8] Source code available at `https://github.com/okfn/opendatacensus/`, live website at `http://census.okfn.org/`

[9] Source code available at `https://github.com/okfn/opendataindex/`, live website at `http://index.okfn.org/`

[10] Adaptation of the "Datasets" section available at `http://index.okfn.org/methodology/`

[11] It may be subject of regulation under some jurisdiction: in this case, the data will not likely be made publicly available and openly licensed

[12] Adaptation of the "Scoring" section available at `http://index.okfn.org/methodology/`

can find it online, then it's digital, even if it is just a scan of the paper the dataset is on

- *Is the data available online?* 15 points
  This question addresses whether the data is available online from an official source. If it's on the Internet and you can access it, it's online. Notice that if the organization emailed you the dataset but didn't upload it to any webpage, it is not to be considered available online
- *Is the data machine-readable?* 15 points
  Data is machine-readable if it is in a format that can be easily structured by a computer. The appropriate machine-readable format may vary by type of data: as a rule of thumb XLS, CSV, JSON and XML files are machine readable while HTML, PDF, DOC, GIF, JPEG, PPT are not machine readable
- *Is it available in bulk?* 15 points
  Data is available in bulk if the whole dataset can be downloaded easily. It is considered non-bulk if applicants are limited to getting parts of the dataset through an online interface (querying a web form and retrieving a few results at a time)
- *Is the data provided on a timely and up to date basis?* 15 points
  This question addresses whether the data is up to date and timely - or long delayed. For example, for election data that it is made available immediately or soon after the election or if it is only available many years later. In case there are no timestamps attached to the data, it might be most fair to mark it not timely or up-to-date

- The following questions examine the legal status of openness:
  - *Is the data publicly available?* 10 points
    This question addresses whether the data is "public". This does not require it to be freely available, but does require that someone outside of the National Go Organization can access it in some form (examples include if the data is available for purchase, if it exists as a PDF on a website that you can access, if you can get it in paper form - then it is public). If a formal and justified request or similar is needed to access the data, it is not considered public
  - *Is the data available for free?* 15 points
    This question addresses whether the data is available for free or if there is a charge strictly related to the data request. If the applicant have to pay a membership fee to the organization in order to get the data, it is still considered free
  - *Is the data openly licensed?* 20 points
    This mean the data term and conditions follows The Open Definition[13], which stipulates that in order for data to be open, it needs to be free to use, reuse or to redistribute (subject at most to attribution or share alike requirements). It is vital that a licence is available (if there is no licence, the data is not openly licensed).

3.3 Assessment and quality review process

The evaluation steps will be similar to the ones taken by Open Knowledge, but they will be adapted and sized to suite the Go scenario. The main actors in the evaluation process will be:

1. *Contributors*: any person submit information about the availability of one of the key datasets in their Place
2. *Horizontal reviewers*: people with experience and insight into (open) data in their countries or region, who will verify the results and make sure they are accurate
3. *Expert (or vertical) reviewers*: volunteers that make a review thematically across several countries
4. *Go Global Open Data Index Committee*, acting as the Open Knowledge in Global Open Data Index: it appoints both horizontal and vertical reviewers if available or find a solution if they are missing

4 Expected Results and Conclusions

How much data is actually being released? What kind of data is it, and in what format is it published? Which countries are the most advanced and which are lagging behind in relation to Open Data?

All these questions are important to answer if we want to understand the state of Open Data in the world of the game of Go.

As a benchmarking tool, the Go Global Open Data Index will help to answer these questions. It will allow people to compare the state of Open Data between countries and to measure progress year upon

[13] `http://opendefinition.org`

year. This will encourage National Go Organizations, Go clubs and Go players to work towards improving the quality and increasing the quantity of Open Data.

Key possible effects by opening and re-using Open Data in the context of the game of Go can be various[14]:

- For the National and International Go Organizations:
    1. *efficiency*: it can help increase management effectiveness and efficiency in operations - for instance, putting data and information online helps save service time
    2. *transparency and accountability*: there is a direct link between Open Data and transparency - the more Open Data there is online, the more factual transparency there is. This, of course, affects large number of actors: data activists, journalists, organization, organization's officials, players and many other can benefit from an increased transparency directly in their actions
    3. *balance of knowledge*: having effectively implemented Open Data policies and engaged other stakeholders into re-using data whenever possible, the asymmetry of knowledge between organizations and their officials in one side and go players in the other may diminish
- For the Go Community:
    1. *increased inclusion, participation, engagement and empowerment*: it is a widespread presumption, that empowering a community with information may strengthen their capacity to participate in a decision-making process. The main challenge is to empower players and journalists so that they can re-use the data to enable better-informed participation
    2. *access to information*: Go communities can have an easy way to learn about their and other countries
    3. *support decision-making capabilities*: with Open Data being presented in an understandable manner, Go communities can build up their own understandings and interpret actions of key issues
- For the whole Go Ecosystem:
    1. *new services*: it stimulates creation of new opportunities to provide useful (innovative) services
    2. *international visibility*: with more Open Data available, it is easier to obtain visibility and let stakeholders have access to information related to the game of Go
    3. *better informed monitoring*: putting players and organizations in a better position to monitor actions and provide with new tools to engage players

[14] Brief summary and adaptation of K. Granickas, Understanding the impact of releasing and re-using open government data, European Public Sector Information Platform, Topic Report No. 8/2013 available at `http://www.epsiplatform.eu/sites/default/files/2013-08-Open_Data_Impact.pdf`