

DATA-POP
ALLIANCE

Data-Pop Alliance and UNDP

Measuring the Unmeasured - SDG Tier III Indicators A PILOT FOR THE GOVERNMENT OF BOTSWANA Final Report

Fabio Celli and Micol Stock

With supervision from Bruno Lepri and Emmanuel Letouzé

1. Introduction	2
1.1 Context and Objectives	2
1.2 Big Data to measure Indicator 16.6.2	3
2. Prototype	3
2.1 Project phases	4
2.2 Methodology and Data Selection	4
2.3 Technical Details and System Architecture	5
2.4 Limitations	9
2.5 Privacy and Intellectual Property	9
3. Results and Evaluation	9
4. Recommendations and Opportunities for Improvement and Scaling	10
5. Conclusions	11

1. Introduction

This project is part of a broader collaboration between Data-Pop Alliance and various Country Offices and Regional Hubs of the United Nations Development Programme (UNDP) to improve the measurement and monitoring of SDG Tier III Indicators in several countries in Eastern Europe, North and Sub-Saharan Africa. This has been done through a combination of workshops, consultations, and pilots. In Botswana, it was agreed with the UNDP Country Office to develop a pilot on SDG 16: *“To promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels.”* Specifically, the pilot focused on Key Performance Indicator 16.6.2: *“Percentage of population satisfied with their last experience of public services.”*

The pilot aimed at offering a technical solution to allow Botswana’s institutions to:

1. Monitor the engagement of the country’s population in dialogue with institutions;
2. Ensure a better understanding of the topics that matter most to citizens;
3. Track potential issues related to the delivery of public services.

1.1 Context and Objectives

Located in the southern part of Africa, the Republic of Botswana is an arid landlocked country roughly the size of France and Madagascar, with a population of about 2.3 million, making it one of the most sparsely populated countries in the world. The population density is of 3 people per square kilometer, 10% of whom live in the capital city of Gaborone. With a GNI (Gross National Income) per capita of \$15,534 and a nominal GDP of \$16.636 billion, Botswana is considered to be an Upper Middle Income country. It has one of the highest Human Development Index (HDI) of the continent, similar to Tunisia’s and higher than South Africa’s, a life expectancy at birth of 67.6 years, and has had uninterrupted democratic elections since its independence in 1966.

Botswana has a highway connecting all major cities, and a public transport system (rails, busses) that is crucial for commuting. Unemployment rate is reported in Figure 1 below. In Botswana, there are 117 internet subscriptions per 100 individuals, placing the country among the 20 African countries with highest internet usage.¹ In this context, policy makers could leverage information coming from citizens activity online as a new data source to design their recommendations.

¹ <https://africaopendata.org/dataset/africa-2018-population-and-internet-users-statistics/resource/3999593a-277f-4cf3-a11c-fa4a71d0e072>

	Unemployed	Employed	labour Force	unemployment rate %
BMTHS 2015/16				
Male	68,535	352,872	421,407	16,3
Female	80,765	341,877	422,642	19,1
Aggregate Age 15 years upwards	149,300	694,749	844,049	17.7
BMTHS 2015/16				
Male	67,187	350,354	417,541	16.1
Female	79,912	340,547	420,459	19.0
Aggregate Age 18 years upwards	147,101	690,901	838,002	17.6

Figure 1: Botswana Labor statistics
Source: <http://www.statsbots.org.bw/labour>

1.2 Big Data to measure Indicator 16.6.2

The SDG Indicator 16.6.2 “Percentage of population satisfied with their last experience of public services” is considered a Tier III Indicator, meaning that “no internationally established methodology or standards are yet available for the indicator, but methodology/standards are being (or will be) developed or tested.”

In order to support the analysis of the population’s level of satisfaction with public services in Botswana, this pilot was designed to look at social media as a source of information. Through the proposed methodology, comments on social media can be used to retrieve data on issues of interest for Botswana’s government, such as tracking citizens’ concerns and hot topics, receiving feedback on public services and more.

2. Prototype

It was decided to develop a prototype social media monitoring platform, called **BotswAnalytics**, to measure and visualize satisfaction, engagement, and topics of interest.

Making use of publicly available data from Facebook, the prototype is designed to collect and analyze citizens’ comments, extract the matters that are discussed, and visualize all the information in real time. In this way we provide the government of Botswana with a tool for understanding the level of satisfaction of people using public service, and even elicit responses on specific topics from online surveys on the official Facebook pages. This method saves money and time with respect to traditional surveys, especially in sparsely populated and rural areas. Additionally, the results reported on the platform can be used for discussing and improving policies related to

areas such as urban planning or employment. The tool always guarantees privacy as no personal information is ever disclosed.

In particular, the government of Botswana expressed the wish to monitor comments in the Official Botswana government Facebook page and in the one of the Ministry of Employment, Labour, Productivity and Skills Development.

2.1 Project phases

The project was structured in three phases, described below:

- I. **Requirements:** The Government of Botswana proposed a set of topics to measure, such as tracking waiting time, feedback on prices and general satisfaction of the users with public transport. Following a data search in social media (Facebook) we assessed the impossibility to retrieve these kind of information at the desired level of granularity. Therefore, we improved the project goals in two directions: first, we extended the range of topics to an open information extraction task; second, we coordinated with the social media management of Botswana's government in order to post polls on the Facebook page on issues of interest. In this way, we were able to retrieve a good amount of data on the desired topics with the analytics tool.
- II. **Development:** We developed the analytics tool, named *Botswanalitycs*, in .php, a common language that runs on every server and could easily be installed. We also provided a user's manual with the system's technical details regarding installation instructions and data retrieval. The tool will be installed on two servers, one for UNDP, and one for the government of Botswana.
- III. **Evaluation:** A good analytics solution constructs a framework for collecting, analyzing and utilizing data to improve the efficiency of specific processes. In our case, the goal was to understand public opinion by looking at the official Facebook page of the Botswana Government.

2.2 Methodology and Data Selection

In order to build a useful platform, we retrieved or computed the following data:

1. Percentage of audience engaged on Facebook, comments and posts share rate;
2. Topics extracted from comments in the monitored networks;
3. Percentage of languages (English, Setswana, Portuguese, French) in the comments.

We extracted topics from the retrieved data and displayed the results in wordclouds (see Figure 2). We also measured engagement on the Facebook page and automatically detected the different languages among the Facebook comments.

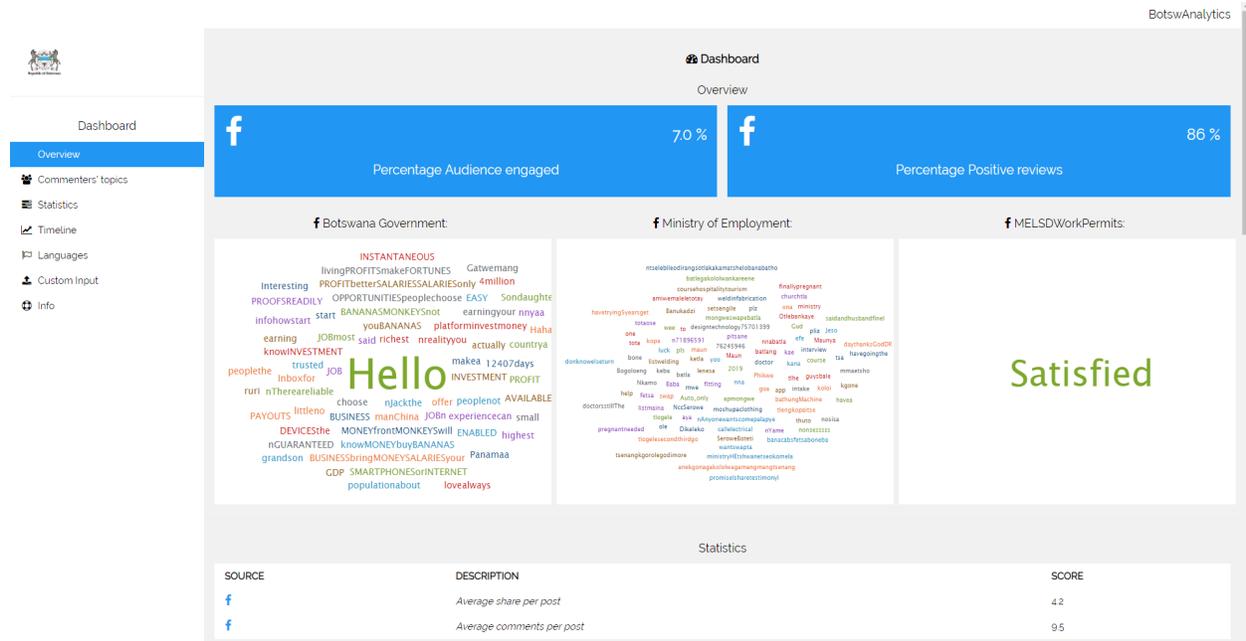


Figure 2: An example of the interface with a menu bar on the left, two overview boxes on the top, three wordclouds in the middle and a statistics table on the bottom.

2.3 Technical Details and System Architecture

Botswanalytics is a Social media monitoring prototype designed for analyzing three specific sources, indicated by the Government of Botswana:

- <https://it-it.facebook.com/BotswanaGovernment/posts>. This is the official page of the Botswana Government on Facebook, where people interact by writing about topics popular among the public
- <https://it-it.facebook.com/pg/Ministry-of-Employment-Labour-Productivity-and-Skills-Development-BW-1328564953862950/posts>. The official page of the Botswana Ministry of Employment, Labour, Productivity and Skills Development
- <https://it-it.facebook.com/melsdservices/posts>. A page dedicated to discussions about services of work permits, delivered by the Botswana Ministry of Employment, Labour, Productivity and Skills Development

While it is not possible to add new sources to the interface, one can upload custom text data obtained online or offline and display it as a wordcloud summary.

The system requirements are as follows:

1. Installation on a server with any operative system that supports *php*
2. *php* version 5 or above installed
3. A domain on a server (the system is accessed via *http* with a browser)
4. No restrictions on file writing on the server

To install *BotswAnalytics* the following steps are needed:

1. Unzip the files distributed in the *.zip* folder
2. Copy the files into a folder on your server, accessible via *http* protocol as a domain
3. Access the folder with a browser as <http://domain.xyz/folder/> or <http://domain.xyz>

BotswAnalytics can be accessed online. The system has been tested on the following browsers: Chrome, Firefox, Edge and Explorer. It is worth mentioning that it can take some minutes to load, as it retrieves data from its sources in real time. The platform contains a menu bar on the left column and a set of boxes in the central column. The menu can be used for jumping from one section or box to another.

The system interface is responsive. This means that it adapts automatically to the device in use, be it a desktop computer or a mobile phone.

Furthermore, there is an overview section that contains two boxes: the percentage of audience engaged on *Facebook.com*.

The percentage of Audience engaged on Facebook.com is computed (calculated) as:

$$\left(\frac{\textit{following}}{\textit{talking about}} \right) * 100$$

The percentage of Positive Reviews on Facebook.com is computed as:

$$\left(\frac{\textit{positive ratings}}{\textit{total ratings}} \right) * 100$$

The following boxes contain **wordclouds** that display a visual summary of the text extracted: the larger the word font, the higher the frequency and the importance of the

word in the text extracted. Words are presented out of context and the system selects the most meaningful ones to represent a conversation. There are three wordclouds:

1. One wordcloud represents topics in Facebook comments on the Botswana government official page
2. One wordcloud represents topics in Facebook comments on the Botswana Ministry of Labor page
3. One wordcloud represents topics in Facebook comments on the Botswana Ministry of Labor services for work permits

Although the sources are different, all the wordclouds are processed in the same way: each one contains topics automatically extracted from a sample of text retrieved from the users' comments on the Facebook source page. The system removes all the English and Setswana stop words, such as prepositions and articles, from this text and then extracts topics expressed as single words. It keeps the 50 most frequent words in order to have the clearest visualization possible. Note that the amount of words displayed in this cloud can depend on the screen resolution.

The **statistics table** reports a set of metrics extracted from the Facebook source page: average share per post and average comments per post are listed. These numbers are reported as raw counts (not as percentages).

The **timeline box** allows the visualization of the engagement percentage over time. The span of time is set on the last six visits to the system. As the system is designed to retrieve new data each time it is accessed, the timeline will display information updated to reflect this feature. The data is stored in the file "*time.txt*" on a daily basis at minimum.

The **languages box** contains a bar plot representing the percentage of the amount of text in English and Setswana retrieved from Facebook comments. The languages are automatically identified with a specific algorithm based on the recognition of frequent words, and extendable dictionaries. The algorithm reports the percentage of languages identified in the sample of text examined.

The **custom input box** allows users to upload a textual file (in *.txt* format) and obtain a wordcloud. Note that when the file is uploaded the page must be reloaded, which can take some minutes. This tool is useful for the collection of offline surveys, that can be summarized and visualized in wordcloud format, like for the other sources, at a later stage.

The system architecture is depicted in Figure 3 below:

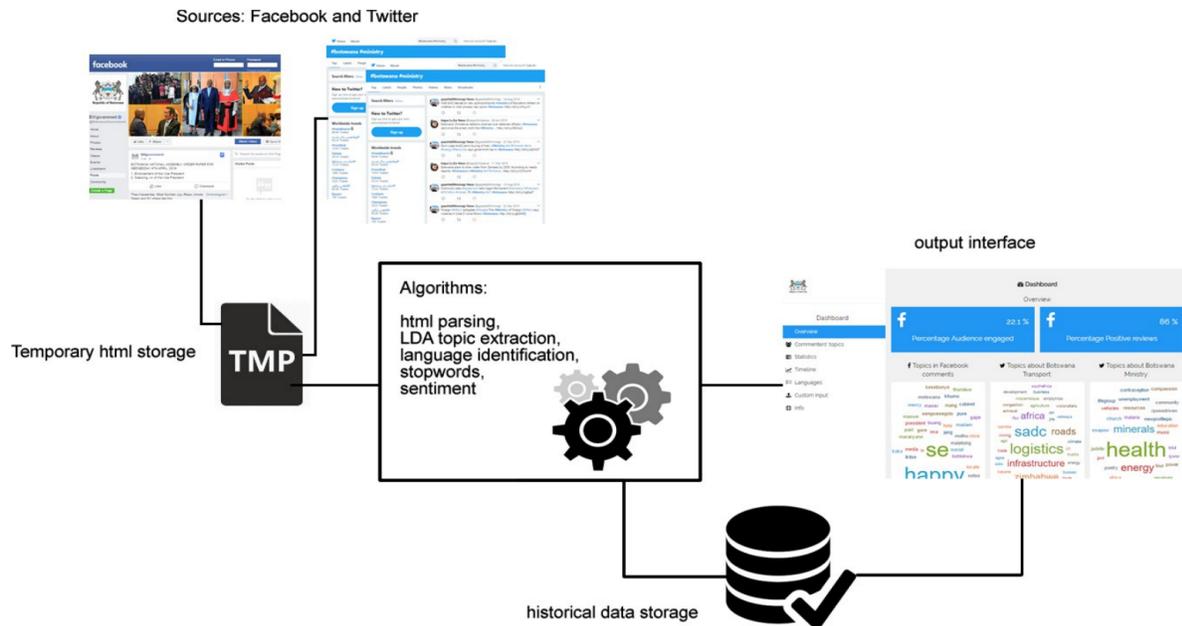


Figure 3:
System architecture

The pages retrieved are stored temporarily into an *html* file and parsed by two different functions, one for the extraction of content from Facebook, in particular these functions extract:

1. page type
2. review ratings (stars)
3. engagement score
4. followers count
5. count of “People talking about” in Facebook
6. counts of comments, likes, shares per post
7. texts of comments

The texts from Facebook comments posts are processed with the following functions:

1. language identification
2. stop words removal (English and Setswana, even using regular expression to improve the coverage)
3. topic extraction
4. sentiment extraction

The text obtained with this process is displayed on the wordclouds in the output interface. The result of the language identification is displayed in the **language barchart**; the percentage of people engaged and the percentage of positive-sentiment reviews are displayed in the **overview box**.

There is a **database** for the storage of the historical data: this is used only for the timeline, which, as mentioned above, displays the 6 more recent updates.

2.4 Limitations

The text is retrieved automatically from Facebook public pages by means of a specific function to parse the Facebook pages from the HTML. Facebook has strict constraints limiting the number of posts publicly available, that depend on the specific privacy settings of the Facebook users. In particular, Facebook policies do not allow the retrieval of posts of the users that are privacy-protected.

Another limitation imposed by Facebook policies yields different results if the page is accessed by a logged human or a system/bot. In particular, a system can retrieve a limited number of comments per post, and these comments are selected automatically by Facebook's algorithm, for which parameters are not disclosed.

2.5 Privacy and Intellectual Property

BotswAnalytics is compliant with Facebook policies, guaranteeing it displays only information that is explicitly made public by the users and is available from public pages. The system does not store personal data, except for temporary processing purposes, and does not present any identifier that could be used for tracking a physical person.

The property of the code and the system, except where otherwise noted, is exclusive to Fabio Celli, 2018 and is provided under Creative Commons License (CC BY-NC-SA). For more information please contact: fabio.celli@live.it.

3. Results and Evaluation

There are no standard procedures and metrics to evaluate an analytics tool. We have designed the following procedure: we measured a baseline of the audience engagement in the official Botswana Government Facebook page. The graph in Figure 4 below represents the engagement curve for Botswana's Government official Facebook page from March to November 2018. The graph shows that the baseline audience engagement is 5% of the page followers, and the peaks correspond to specific events, such as the installment of the new government. These metrics will be useful as a baseline to measure the impact of future campaigns, that can be compared to the present data.

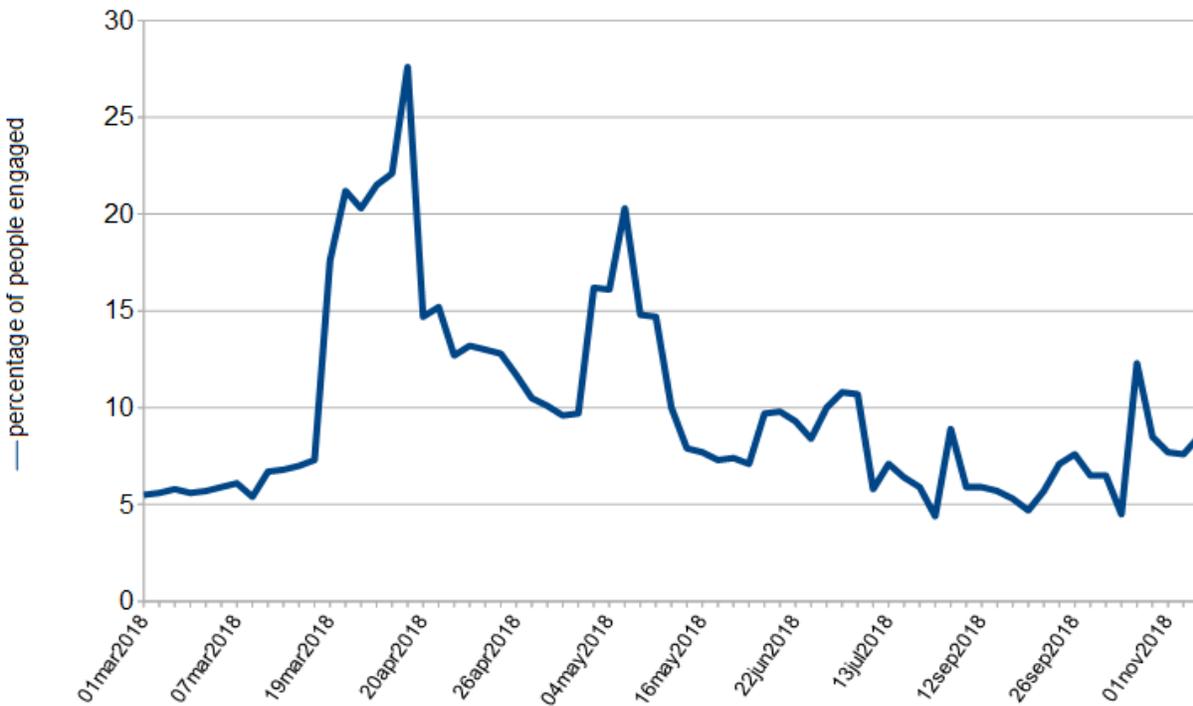


Figure 4: Engagement curve in the Botswana Government Facebook page from March 1st 2018 to November 4th, 2018

In order to conduct a qualitative and quantitative evaluation, government offices plan to run surveys through the official Facebook page and collect the responses that will be visualized in the platform’s wordclouds. The evaluation phase could be expanded when the system will be running on two government’s servers and upon full implementation of a survey.

4. Recommendations and Opportunities for Improvement and Scaling

The proposed system and methodology can easily be applied to other government offices, settings and countries in order to contribute to measure SDG Tier III Indicator 16.6.2. The technological solution is indeed scalable since it can be implemented in any country that has an official Facebook page to monitor.

The adaptation to new scenarios implies changes in the code, which is open-source.

To overcome the limitations imposed by Facebook, a recommendation would be to redirect the traffic of people wishing to compile the surveys to a landing page outside Facebook. In this way, it would be possible to collect data externally and at a later stage upload it into the system by means of the custom input wordcloud, that allows a visual

summary of the textual data. Moreover, we would recommend organizing public events, hackathons and targeted workshops in order to increase community engagement. This will help build capacities and incentives to foster citizens' online activity as well as contribute to more evidence-based policy making.

5. Conclusions

This pilot aimed at demonstrating how Big Data could be used to estimate citizens' satisfaction with public services, considered to be a Tier III Indicator among the SDGs. There are strong incentives to implement new methodologies to produce high quality statistics. The ambitious 2030 Agenda set forth by the United Nations in 2015 and endorsed across the globe by nearly every member nation will be all but impossible to achieve without greater availability of quality, timely data which governments may leverage to make evidence-based decisions, and for citizens to hold them accountable.

In this context, BotswAnalytics, the system designed specifically for Botswana, is a concrete example of how Big Data can help measure progress towards the SDGs. Although there were a few challenges in the actual implementation of the tool, mainly due to time and administrative constraints, we still see the potential to scale up the project and expand the analysis to other government offices.