Data Use Case Study: Opening public data to improve government transparency in Nepal

Prakash works to improve the availability and accessibility of government budget and spending data.

Prakash Neupane is the ambassador of Open Knowledge Foundation's Nepal (OKFN) Local Group. Embracing their role as a data intermediary, the group develops technical solutions to open up data around social issues, such as government accountability.

"Civil society should be given every opportunity to access and understand budget information better". Prakash Neupane, Ambassador, Open Knowledge Foundation Nepal
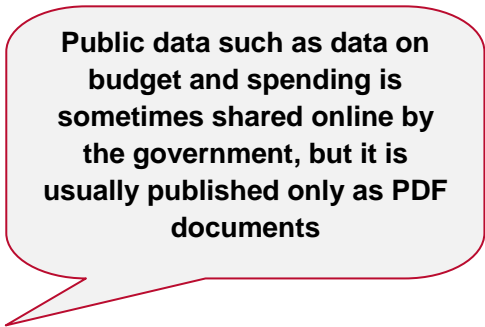
The budget is a government's most important tool for translating policies, political commitments and goals into decisions about how to finance the needs of a country. The budget outlines the proposed revenues and spending for the financial year, and the allocations to specific departments, sectors, localities and projects. As such, it has wide-ranging implications for citizens, impacting both directly and indirectly on their well-being and prospects. Civil society groups and the media have an important role to play in tracking that the budget is appropriately formulated, executed and overseen. By subjecting the budget to debate and scrutiny at all stages of the budget cycle, civil society can support the effective allocation of resources and monitor for instances of inefficient spending. In order to do this effectively, there needs to be timely, accurate and comprehensive budget information available to civil society in a format that can be easily used for

analysis and comparison, which they can convert into meaningful messages for citizens. In Nepal, until recently, there was no government budget data available in this format.

Prakash Neupane is the ambassador of Open Knowledge Foundation's Nepal (OKFN) Local Group. The OKFN Local Group is a party of tech enthusiasts, comprised largely of university students, who meet on a regular basis to engage with issues of interest around open content and open data[1] – whether it is open culture, open science, open statistics, open coding or open finance. The Group meet to learn, innovate, share, support and explore ways to build their data literacy skills around finding technical solutions to topics of particular interest to them. It was at one of these Local Group meetings that the issue of the government budget was raised. "Budget data is something that matters very much to people because any adjustment in it has a direct impact on individual household incomes and living standards". Prakash explained that members of the OKFN Local Group were aware that budget information did not exist in a form that could be easily used, but felt strongly that "civil society should be given every opportunity to access and understand budget information better". With this in mind, the Group decided to see what they could do to help.

The Group settled upon a plan to transform the government budget and spending data from one locality into an open format. The Kathmandu Metropolitan area was chosen because it has one of the largest revenue collection agencies in Nepal, and one of the largest recipients of the government budget. The first step for the Group was to find out how to obtain details about the local government budget and spending information. They approached the Kathmandu Metropolitan office and found out that the data was publically available online via their official website. The second step was to find out in what format this information was made available. Here they discovered that the data was available only as PDF files. "In Nepal, public data such as data on budget and spending is sometimes shared online by the government, but it is usually published only as PDF documents". Prakash explained that "government authorities often mistakenly perceive data as made 'open' when they manage to publish it in PDF files on their official websites. This is not a usable format, and prevents analysis of the data contained in the PDF". It did not take the Group long to discover that there were a range of other difficulties in accessing, using and analysing the budget information. As the data was not in open format it was not licensed in a way that enabled the user to know if they could re-use or redistribute it, and much of the information provided was headline figures rather than data in disaggregated detail. Prakash believes that "for budget data to be useful for analysis it needs to be opened, detailed and disaggregated, and for budget data to make real difference in the lives of people, access to smaller data from district, ward and village level should be ensured".

> **Public data such as data on budget and spending is sometimes shared online by the government, but it is usually published only as PDF documents**

---

[1] Open data is data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and share alike

In mid-2013, OKFN hosted a two-day event to liberate the Kathmandu Metropolitan budget data, calling it a 'Data Spending Party'. This event brought together OKFN Local Group members with coders, journalists and students to extract the city spending data from the online PDF documents. This involved using computer software to 'scrape' the data and transform it into an open format. According to Prakash, "We had to invest a lot of time in scraping the spending statistics from the PDF documents as there were several challenges". The first challenge was identifying a method for doing this – having researched a range of products, the Group chose 'tabula' software to extract data into MS Excel and then used 'Google Spreadsheets' to clean it. The sheer quantity of data to be scraped posed the second challenge, with over 20 pages of documentation needing to be scraped. Thirdly, "much of the data was in a form of Nepali font that did not use the Unicode standard, making it difficult to convert it in English". The Group felt it was important for the spending data to be translated into English so that it could reach a wider audience, and also because unless a user's computer has the Nepali font installed on it, the computer is not able to read the Nepali script. A fourth challenge was that the PDFs from the Kathmandu Municipality used inconsistent fonts, which distorted the order of the dataset when extracted using tabula. Prakash shared that, "in total, it took around 10 people who worked over 36 hours to scrape and clean the data from the PDF documents, getting the data into a format that could allow analysis and redistribution".

Once the Kathmandu Municipality budget and spending data has been opened, Prakash and the Local Group were able to sort and filter year-on-year data using MS Excel, enabling them to make multiple analyses. "The analysis we were now able to undertake revealed some interesting trends in the data – ones that it would have been difficult for us to see before". For example, the Group noticed that Kathmandu Municipality's spending increased significantly over time in the areas of infrastructure development and employee salaries. Recognising the value of this kind of information, the Group decided to present their findings to a wider audience, to demonstrate the potential of the data. "We have tried to make the analysis visually engaging through the use of graphs, charts and visualization tools. This has allowed us to share the budget and spending information more widely, for example with journalists, civil society organisations and with the suppliers of the data themselves". According to Prakash, the Kathmandu Municipality office has been very appreciative of the analysis that the Group has done using the opened data. "The officials have expressed an interest in publishing their future spending data in open format – we consider this a great success. We hope that we will be able to collaborate with the Kathmandu Municipality again in future, and possibly help other areas to publish their budget and spending data in open format".

Boosted by the success of their first public data liberation, the Local Group decided to attempt a second data liberation project in early 2014, but this time attempting to open up a much larger and potentially more significant dataset – the Nepali government's national budget data. Having learned techniques and lessons from the challenges experienced during the first attempt, the second attempt ran more smoothly. The Group found that the budget data from the recent fiscal year (2013/14) was easy to locate in a PDF document published on the Ministry of Finance website. The data scraping event was hosted to coincide with Open Data Day 2014. Prakash explained that "although the data scraping process was still cumbersome, this time it was less

challenging to convert the font into English. This was partly because the PDF document had been published using Unicode and so the font was less problematic to scrape". A team of eight people managed to scrape the budget in just under six hours. The data was then published on the Open Spending website.

Prakash sees interesting possibilities for the use of the newly opened budget data, "Now that it is open it can be joined up with other datasets for cross analysis to be used in research, reports or advocacy and also as a basis for developing apps that use budget data to help people make decisions based on statistical evidence". He believes that greater access to budget data will create more opportunities for civil society to participate in budget formulation and to monitor if it is being spent as stated. "It has also opened up avenues for public discourse that can help identify which sectors receive just small budgets, and which receive more funding than what is required". In his view, both governments and donors benefit from this openness; feedback from civil society enables the government to make budget more responsive to citizens' needs, and donors can make more effective decisions about where to place their funding.

> **Now that it is open it can be joined up with other datasets for cross analysis to be used in research, reports or advocacy and also as a basis for developing apps that use budget data to help people make decisions based on statistical evidence**

Prakash believes that in an ideal world, this sort of budget data would already be available in an open format. "If this data was released by government in open format we would, instead of spending long hours scraping data, be using our time to carry out analysis and make the findings interesting for audiences to read". The efforts that technical intermediaries, such as the OKFN Local Group, are making to improve the availability and accessibility of data are key pre-requisites for the use of such data by civil society. Prakash hopes that "by opening up public datasets like budget data we can make it possible for civil society to use the data for monitoring and decision-making, and ensure that the government budget achieves its full potential"