

Open Data is only the Beginning: Civil Society Organizations as Public Data Analysts

Transparency and open data in the current discourse are often portrayed as useful as ends in themselves. However, data must be turned into understandable information to impact decision-makers and influence public opinion. Analysis of data and then effective communication of that analysis by intermediary organizations is a crucial but under-discussed aspect in the open data ecosystem.

Civil society organizations focused on governmental accountability are often ideally positioned to take play this critical role. Once these actors analyze and transform data into useful information, they would need to communicate key findings effectively to their target audiences in order to advance an advocacy or other political agenda.

Drawing on lessons from the efforts of the Coalition for Democracy and Development in Ghana (CDD) around the December 2012 elections, this paper will consider the pragmatic ways open data --and the necessary analysis-- could move from an academic ideal to a political reality with the help of civil society organizations (CSOs).

Background on Ghana and Election Data

Ghana is a successful, stable democracy that had a tense but ultimately successful democratic transfer of power during the 2008 elections. In the advent of the 2012 polls, Ghanaians and international observers were yet again concerned about the prospect of a clean elections. There, as in other countries, civil society organizations were gearing up to monitor the election. Citizen oversight of elections is a proven method of ensuring that an election is free and fair, and that the results as announced reflect intent of the people and the actual votes cast.

NDI has worked with scores of civil society organizations and coalitions over the years to monitor elections. A core innovation, initially developed for the 1997 elections in the Philippines, was the concept of a “partial vote tabulation” (PVT). Trained volunteer observers are deployed to a statistically representative sample of polling stations across the country, monitoring these polls as to the conduct of the voting there throughout the entire day. Observers report on the conduct of setup, voting, and counting during regular intervals throughout the day, sending structured SMS with this data to a central data center run by the monitoring organizations. After the day is complete, these observers also report the tabulated election results directly from their polling station. Statisticians and analysts at the data center rapidly crunch these numbers, and are able to draw representative conclusions

about the conduct of the election, and, more importantly in many countries, the actual results.

In Ghana NDI partnered with CDD and its network of observers, CODEO, the Coalition of Domestic Election Observers, to monitor this year's critical and hotly contested polls. CODEO, the largest monitoring coalition in Ghana, has been observing elections since 2000, and is among the most experienced and sophisticated partners with whom NDI works. They are backed by a team of social scientists and statisticians with the ability to do sophisticated analysis of the mountain of data produced by their observers throughout the course of election day.

CDD's effective analysis of this information and creative new methods for communicating it to the public makes this a fascinating example for other civil society organizations looking to take advantage of new data resources and effectively share them with their fellow citizens in support of a specific political goal.

Traditionally, analysis of the massive amount of observer data (in some countries, upward of 60,000 SMS during election day) has been released via press conferences, formal statements, and (months later) long-form reports comprising many pages of densely written text.

However, the audience for such information has always been limited; while journalists may produce short pieces based on information shared during the course of a press conference, the post-election analysis only reached a key few -- the most critical of decision-makers: party leadership and the incumbent administration; the embassies of interested countries; regional and international observation groups, and other such political elites.

Open Data and Civil Society Organizations

To create the initial monitoring plan and inform the analysis, CDD integrated information from various other open data sources. Most important was that of the Ghana Election Commission. This data may have been freely available, but it was not easily accessible: NDI had to scrape information from the PDFs provided by the Commission to get complete lists of polling stations and to capture past election results. Some information, such as polling-station level election results from past cycles, was never publicly released.

Other critical information to aid analysis came from demographic information from the World Bank. Gender, ethnicity and socioeconomic status information can be powerfully

correlated with the information gathered by the organization. In the 2009 elections in Afghanistan, a major data analysis project included information on the recorded number of security incidents in a region; which correlated to polling stations that went overwhelmingly - one might say implausibly - towards the incumbent.

In Ghana, it is worth noting that the data generated by the election monitoring effort was not itself directly released in an open-data format. CDD, and organizations like it, are often loath to release the complete datasets of their election monitoring results. There are a number of reasons for it: such data is inevitably messy and will contain mistakes or gaps that could be used to discredit the process; if the data is not sufficiently anonymized, individuals who reported at particular polling locations could be targeted for retribution; different data points are used in different ways that are not obvious to outsiders; hostile governments could determine the strategy used and attempt to undermine future efforts.

Open Data Requires Analysis - And CSOs Can Deliver

Open data is a raw material that is not very useful on its own for the vast majority of citizens. As more countries embrace principles of open data and start sharing government data in machine-readable and more standardized formats, this will create supplies of information that will require parsing and analysis by intermediating actors to be useful for a general public. There are a number of different types of organizations that are well-placed to serve this translation role; journalists and academic experts, for instance. However, civil society organizations able to embrace this new source of information are ideally placed to understand the source information and then communicate it effectively the public - as well as use such data streams to advance specific political or policy goals.

The product of an organization like CDD is not data, but rather analysis. Data is an input to the organization; their political knowledge and statistical expertise come together to take this heap of unfiltered election observer information and create digestible knowledge out of it that meaningfully tells a story - in this case, about the conduct of an election. As such, they are already poised to work with new sources of information that they themselves have not collected.

Once You've Got the Goods, Analysis Needs to be Communicated

To be effective in this intermediary role, the processed information needs not only be turned into useful analysis but also effectively shared with public audiences. It is the tree-falling-in-the-forest problem --the best information in the world is of no use if it not

communicated with target audiences. Once an organization has useful information gathered from open data sources or its own research and analyzed it in a compelling fashion, it needs to be shared. Transparency is a necessary but not sufficient goal: data, for an end-user audience, needs to be readily consumable to be useful.

There are many different potential groups that could be target audiences for data analysis provided by CSOs. Audiences differ in two critical ways: their level of sophistication and the most effective way to reach them. There is no one-size-fits-all answer for how to communicate with all people in the same way at the same time. With apologies to Marshall McLuhan, the medium may not be the message, but it does dictate who is waiting on the other side, which, in turn, this provides opportunities for targeting the level of analysis and amount of context to the typical recipients. At the same time, different methods of distribution are suited for different types and styles of information.

This year CDD has found new ways to release analysis of its data to make it useful and accessible to new audiences, breaking out from the small circle of those who would have been likely to pick up the information from traditional print articles.

CDD distributed information via three new channels during the 2012 election:

Radio

As in many countries, radio is a widely accessible and inexpensive medium to reach a mass audience in Ghana. There is no communications medium that can come close to radio for reach from the wealthy in Accra to poor communities in northern Ghana. Newspapers have limited circulation that is largely confined to urban areas; internet access is far from universal. Everyone listens to the radio.

CDD decided to create a series of brief radio programs for community radio stations across Ghana, particularly in the more rural communities. As such, audiences would not be highly educated or sophisticated, and messaging needed to be appropriately pitched. A series of programs begin with the basics of “what is election monitoring” explaining the power of a statistically-based observation mission. Another went into depth on the challenges faced during the biometric registration process in the country; others discussed the most common issues observers found during the pre-election period.

On election day, CDD produced a mid-day program that drew upon their data collected through the process so far to brief listeners on what was taking place as information came in. Others took the high-level analysis of the results of the election and evaluated the process of the entire day. With the election complete, CDD is creating another series that

highlights critical elements of the data that the analysis has been able to tease out in short, engaging audio clips distribute to radio stations.

Social Media

Most Ghanaians are not on Facebook, but those who are form a critical audience. Political leaders, elites, young people, and opinion-shapers are disproportionately represented. Those online are likely to have a somewhat higher level of sophistication and knowledge about the election.

CDD aggressively moved into Facebook to share the election analysis with these key audiences. On the back of a relatively inexpensive paid advertising strategy, the CODEO Facebook page went from 120 to 9000 followers in about 10 days, and their friends-of-friends encompassed virtually all Ghanaians on Facebook, with a reach of approximately 1.6 million.

Given the nature of the medium, CDD took the analysis and broke it into smaller atomic elements, focusing on one small element at a time that was demonstrated by their data. For example, their pre-election report was a lengthy work of analysis encompassing the totality of the environment in the runup to the election. Using Facebook, the CDD social media team posted highlights such as details on the biometric registration process.

Mapping and Visualization

Maps are all the rage these days; however, many of them exclusively focus on putting up raw data such as incidents or vote totals. As with other types of open data, visual information is more useful if it is filtered through analysis and made more accessible to different, wider audiences. The most compelling innovation that CDD produced with regard to visualization this cycle was the ability to render the quality of the electoral process throughout the day. The raw data of how individual polling stations performed on the various questions evaluating how well the process was executed were aggregated at a constituency level. At a glance, viewers got an easy-to-understand depiction of the relative quality of the electoral process in different regions of Ghana.

Politics Matter

Well-intentioned concepts for open data analysis can wither in a given political environment. Civil society organizations and other intermediary organizations operate within a political context that can get quite uncomfortable at times, especially in emerging democracies where this political space is hotly contested. In Ghana, the election results were extraordinarily close, and are at the time of writing, were being contested before

Ghana's Supreme Court by the losing party. Particularly around elections the stakes are incredibly high, and organizations may be inclined to be cautious.

One way to preempt such challenging situations in the heat of the moment is to have set up key points of analysis -- the story line that the organization tries to convey-- and distribution channels in advance before results have been calculated. By setting up a system through which open data feeds or other elements are processed automatically and displayed according to pre-determined ways that support the key messages an organization wants to get across, an organization is able to bind itself to a sensible course of action in advance of knowing the outcomes.

Conclusion

Open data and governmental transparency opens a door to knowledge, but is not directly useful in and of itself. Data needs to be processed and analysed to become useful information, and then digestible knowledge must be shared with interested audiences. Civil society organizations are well placed to play this role of translator, guide, and political advocate using such data, enabling citizens to make effective use of the critical information governments share via open data channels. The past success and recent innovations of CDD and the CODEO monitoring organization provide a useful model for other civil society organizations looking to make open data useful to their fellow citizens.

To complete the chain, governments must be transparent, creating open data or permit civil society organizations like CDD to gather their own. Civil society then needs to take responsibility for applying their analytical skills to the collected data, but creating valuable digested information is also not enough. That knowledge then needs to be effectively communicated to the citizens of the country, using effective communications media and appropriate levels of analysis. Intermediary organizations such as NDI have a key role to play in sharing best practices and encouraging organizations to consider the ways in which their analysis tools and communications techniques can evolve to keep up with the state of the art. Working together, these groups can complete the open data lifecycle, fulfilling the promise of governmental transparency by using it to help shape an informed citizenry.