Many Hands Make Light Work

Crowdsourcing geoenables development assistance data

By Karen Richardson, Esri Writer

In just 16 hours, a crowd of volunteers processed 117,000 records for US Agency for International Development (USAID) that showed where credit was made available to entrepreneurs around the world through loan guarantees extended by USAID's Development Credit Authority (DCA).

The crowdsourcing event, held to make international development data more accessible and transparent, was not only a first for the agency but also the first time the approach was used by the US government to open data and use Data.gov as a crowdsourcing platform.

An independent agency, USAID has provided economic, development, and humanitarian assistance around the world in support of the foreign policy goals of the United States for more than 50 years.

Getting by with a Little Help from Friends

Crowdsourcing is a relatively new phenomenon that has evolved significantly since the emergence of Web 2.0. In the words of Jeff Howe, contributing editor at *Wired* magazine, "Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call." While still a nascent movement, increased public participation using crowdsourcing and other new technologies presents a shift in how the US government engages with its citizens and how citizens can participate in their government.

In the context of humanitarian and development, crowdsourcing was used in

response to the 2010 earthquake in Haiti. Over the last two years, as the information landscape has continued evolving, this sector has identified innovative ways to incorporate new data and methods into well-established workflows. The use of crowdsourcing for humanitarian or development interventions has spurred a lively debate about the advantages and disadvantages of this approach. The discussion has included many questions relating to data quality, security, and usability. These issues were carefully addressed in USAID's case study on crowdsourcing open data.

Putting the Crowd to Work

To enhance understanding of the geographic distribution of loans guaranteed by the agency, USAID's GeoCenter worked in cooperation with DCA to identify a global dataset of records to map and make available to the public. Interested individuals, including volunteers from the online technical communities Standby Task Force and GISCorps, structured data on certain USAID economic growth activities and then geocoded the data. The data cleaning and geocoding process, originally estimated to take 60 hours, was completed by the volunteer technical communities 44 hours earlier than anticipated.

In the DCA database, all geographic information is stored in a single field that is not standardized across all records. Sometimes the city is given, sometimes a street address, and sometimes only the first administrative level (Admin1) or state level. The data must be broken out into different fields so it can be mapped at the lowest level of granularity across all records. Once parsed out, the city name can be used to capture Admin1 level information using a gazetteer such as the National Geospatial-Intelligence Agency's (NGA) GEOnet Names Server.

Given the country name and original location data, volunteers mined the data for clues that would allow them to determine the correct Admin1 unit, Admin1 code, and place-name for a populated area. (The Admin1 code is designed to eliminate problems of transliteration between disparate languages.) Because the data was problematic, not all records could be processed, so volunteers flagged some records as bad data.

USAID worked with Data.gov, Esri, and Socrata to develop web applications for the project. Socrata, a developer and provider of open data services, designed a custom application that allowed Data.gov to be used as a platform for data editing and generation, ⟶

↑ Getinet Enyew, who grew up in Ethiopia, left the country when he was younger but always dreamed of going back to be part of his country's development. When he learned a local bank was approving loans for qualified members of the diaspora who returned to start businesses, he jumped at the chance to build an organic farm in Menagesha, Ethiopia. (Photo credit: Morgana Wingard)

considerably extending its original viewing capabilities. With the Socrata application, users could check out as many as 10 records at a time from the database for processing.

Esri created a custom web map application available through ArcGIS Online that allowed users to easily and quickly find administrative names and codes. Because volunteers registered on Data.gov, each geocoded record could be traced back to a specific volunteer. Records were linked to volunteers and USAID staff members could perform spot checks during the crowdsourcing event to look for anomalies in how data was being entered. If a volunteer was incorrectly entering data, that volunteer could be contacted directly or any records they edited could be redacted from the final product if necessary.

Opening Doors with Data

The geocoded data can now be visualized as an Esri story map or shared on ArcGIS Online. Spatially enabling this data lets the private sector explore new areas for collaboration with host countries, researchers, development organizations, and the public. With an appropriate basemap and the addition of other content found on ArcGIS Online, such as world demographic information, organizations and citizens can create, save, and share maps and web applications.

Since the event, USAID has been notified via Twitter that visualizations using the data have been created. One example is OpenSpending, a project operated by the Open Knowledge Foundation (openspending.org/usaid_dca) in Cambridge in the United Kingdom. A visualization on the organization's website displays USAID data from the crowdsourcing project and further enables discussion, analysis, and action related to important development strategies.

Maximizing Impact

USAID anticipates that the release of this data will cause several substantive impacts: better service to entrepreneurs, targeted lending, more insights into opportunities for cooperation across national boundaries, and improved partnerships with other donors.

By creating a map specifically listing available financing, USAID is making it easier for entrepreneurs to see where they could qualify for local financing. In addition, organizations working to help certain groups of entrepreneurs around the world access financing can take advantage of the USAID guarantee map to connect their networks with available financing. While the map does not list bank names or contact information, it provides a contact e-mail address so individuals can connect with local banks via USAID staff.

← USAID staff and volunteers at the USAID Innovation Lab in Washington, DC, during the crowdsourcing event
(Photo credit: USAID)

be easily analyzed across country borders with beneficial results. For example, if the map shows that all loans in one region of a country focus on agricultural projects while loans in a bordering region in another country fund infrastructure projects, this may lead to future collaborations between these adjacent USAID missions.

While USAID and other donors often try to collaborate to maximize impact, there is no overall database of active guarantees offered by all development agencies. By making available the map service layers, other donors can compare or even overlay their guarantee data to identify opportunities for increasing collaboration.

Committed to Open Data

"The US government is committed to opening data and increasing aid transparency. This pilot is an example of this commitment," said Eric Postel, assistant administrator for economic growth, education, and environment at USAID. "By enabling the crowd to help us sort through and clean nonconfidential data, we are able to release information that we never previously thought was possible."

USAID hopes that this initiative will encourage government agencies to move forward with their own crowdsourcing projects. Whether the intent is opening data, increasing engagement, or improving services, agencies must embrace new technologies that can bring citizens closer to their government. "USAID will also continue to explore unique ways to engage the public in our work," said Ben Hubbard, director of USAID's DCA. "Development isn't something that happens overnight. But with increased transparency, we can start working together to solve development challenges in a more efficient—and fun—way."

Visualizing loan data on a map can also change the way USAID field offices (or missions) plan for future guarantees. Often the goal is to make guarantees available in areas outside capital cities or in certain regions of a developing country. By seeing where the loans are concentrated, USAID missions can better determine whether the guarantees are reaching targeted regions. In addition, USAID can overlay other open datasets on the USAID map. For example, by adding a layer of open World Bank data describing access to financial services organized by income group, USAID can quickly see where its activities line up with other financial indicators.

Previously, USAID missions in one country didn't have time to analyze the location of all loan guarantees in surrounding countries. For the first time, USAID loans can

"By enabling the crowd to help us sort through and clean nonconfidential data, we are able to release information that we never previously thought was possible."