A technologist tackles information innovation

And you shall be fishers of data

Also in this section
2 Data visualization in action

OAKLAND

A Solution to the Innocentive Challenge goes global

To a technologist, teaching a man to fish is an awfully inefficient use of time. Faced with such a task, a thoughtful technologist would develop a technology that simplified fishing to the point that it could be performed easily and efficiently by any man, woman or child.

To put it another way - why give a man a fish when you can give the world the fishing pole?

It is with this mindset that I have set about cheerfully extending the requirements for the Economist's Innocentive Challenge. Where the challenge asks Solvers to take a dataset and use it to improve a city, I present a working set of technologies that enable anyone to use any dataset for any number of projects.

My presentation will begin with the Cholera epidemic in Haiti, then move halfway around the world to rural Afghanistan and conclude with an ongoing project that combines 15 datasets to protect 1.5 million people in 14 different cities.

By far, the easiest way to understand these technologies is to explain why, and under what circumstances, I built them.

Haitian Fusion

In late 2010, a Cholera epidemic swept through the earthquake-ravaged country of Haiti. I was contacted in December of that year by a Haitian NGO who wanted to collect information from the hundreds of field hospitals, called Cholera Treatment Centers (CTCs), that had been set up across the country by the government and various NGOs.

We identified three main goals for the project:

- Visualize the locations of the CTCs in an online map, making it easier for relief organizations to deliver supplies;
- Track information from the CTC's on the number of patients they were treating, providing a mechanism for tracking the epidemic;
- Use the location and patient volume data to deliver aid to where it is needed most.

As the plan began to take shape, I began to code. I started with an existing open source web platform named Ushahidi to provide the map visualization and basic search, then extended the platform's information architecture to add per-field security restrictions. This enabled us to make a CTC's location public but restrict access to sensitive information such as the CTC Director's personal cell phone number. I also built a history-tracking system that could record changes to each of the data fields (e.g. number of available hospital beds) over time. After two months of programming I'd reached the limits of what I could do remotely.

I landed in Haiti in mid-February to meet with the Haitian NGOs I had been working with and present the system we'd developed to the Haitian Ministry of Health (MSPP). Our plan was to use the official list of CTCs as the central data set for our project, but after a week of meetings, we had found no less than five "official" CTC lists—each in its own format and containing different locations.

This confusing situation presented a major opportunity: if we could intelligently fuse these datasets, we could create an aggregated map of the CTCs containing more infor-

mation than any of the individual lists. As a result, we identified three core technologies we could use to meet the project's goals:

- Data Fusion: The ability to import and merge all of the available lists without requiring each organization to modify their information collection process.
- Searchable Map: The public information from each dataset visualized on an easy-to-use map with a built-in search system.
- History Tracking: Recording every time a value in any of the datasets changed and using that information to track the epidemic and target aid delivery.

The missing piece was the ability to merge the newly discovered lists. With three days until our presentation to the Ministry, I sat on balcony overlooking Port-au-Prince and built a Data Fusion system. The resulting code was able to understand each group's data format, identify errors and duplicates, and upload the information to the online map. With three hours left before our presentation I uploaded one of the official lists and the system immediately identified a number of errors including one hospital location that was wrong by ten miles. I added a slide about the issues to my presentation and headed towards the National Palace.

I presented the system along with our initial findings and was immediately accused of lying – one of the participants refused to acknowledge that their official list had any mistakes. I knew we had built a powerful set of technologies, but we were under no illusions about our chances of the project going forward. By the time we received word that the Haitian government "had decided to go another direction" we had the technologies up and running in Afghanistan.

Afghani Wells

In the deserts of Afghanistan, a functioning well can be the difference between life and death. USAID, the United States Agency for International Development, has sponsored a project to track the location and status of wells across the country. A colleague asked me to help update the project's well tracking system. The team's existing approach attempted to manage thousands of wells in a single Excel file—it was inefficient, error-prone, and an excellent opportunity to use the technology developed for Cholera tracking in Haiti.

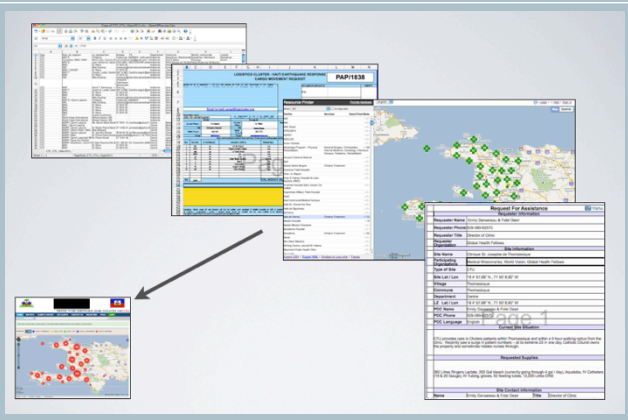
The Data Fusion system checks the well management team's Excel file for errors and uploads it to the online map. Teams of project administrators can search the data and villagers can send notifications that their well is broken using an inte-

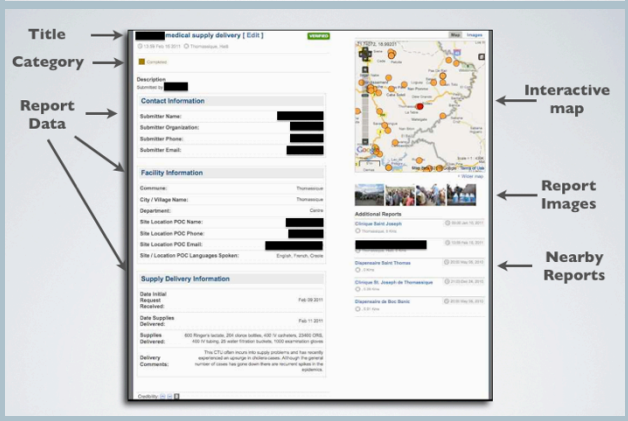
Cholera epidemic

Data visualization in action

Port-au-Prince

Slides from my presentation to the Haitian Ministry of Health





grated phone system. The history tracking system was extended by another colleague to track when wells break and were fixed, enabling the administrators to track which wells were undergoing the most maintenance and how long the local repair crews are taking to respond.

The project has been operational across Afghanistan for the past eight months and is maintained by local Afghans. The Data Fusion system and associated software is operating smoothly—receiving inputs from workers in the field, detecting when errors and mistakes occur, and ensuring that communities throughout Afghanistan are being supported. With one successful deployment of the system in place, I am now applying the Data Fusion system to the data challenges of a community much closer to home.

Californian Quakes

Alameda County covers most of the eastern half of the San Francisco Bay Area. It is home to over 1.5 million people and the most active Disaster Services team in the American Red Cross. For the past several years I have been volunteering with the Red Cross Disaster Assessment (DA) team—the group responsible for maintaining the organization's emergency resources list—a set of fifteen different datasets including private Red Cross data on shelters and supply trailers and public data sets of hospitals, police stations, and fire stations.

Collecting the information is relatively easy, but ensuring that it is available to the Red Cross teams in the middle of a major catastrophe requires a process that is:

- Simple: Explainable to a new volunteer in under 3 minutes at 4am in the middle of an emergency.
- Available: Accessible when the power is out and the Internet is down.
- Maintainable: Easy to keep the information up to date over time.

Our current solution is based on the technology I developed in Haiti and refined in Afghanistan. The DA's official data source is the Excel spreadsheet that is printed and shared with core volunteers. That spreadsheet is then uploaded to an online Ushahidi website using the Data Fusion system. An additional code module developed in conjunction with colleagues at Google's Crisis Response Team enables the same information to be downloaded from the website in Google Earth format. The electronic data files are then distributed to each of the volunteers on USB memory sticks embedded in Red Cross wristbands. All of the instructions for using the various data formats are explained in a six page handout that's mostly pictures.

Disaster data – it's all in the wrist

This approach allows the team to view the information they need through the website if the power is on and the Internet is available, from their memory sticks and laptops if the Internet is down, and from the hard copies if the power is out.

The DA team will perform our second end-to-end exercise of the system on April 15th, and we will have the first batch of binders and wristbands distributed to key members of the Red Cross by the end of the month. From that point forward, management of the system and the information it provides will be controlled entirely by Red Cross volunteers.

Casting a Wide Net

Each of these projects are made possible through the thoughtful creation of reusable technology that has been applied around the world from Haiti, to Afghanistan, to the San Francisco Bay Area. The last year has been quite challenging, and it is just the beginning. I am now working on a number of exciting new projects that use these technologies and anticipate many more opportunities in the future—because when it comes to data visualization projects, I have found that there are plenty of fish in the sea. ■