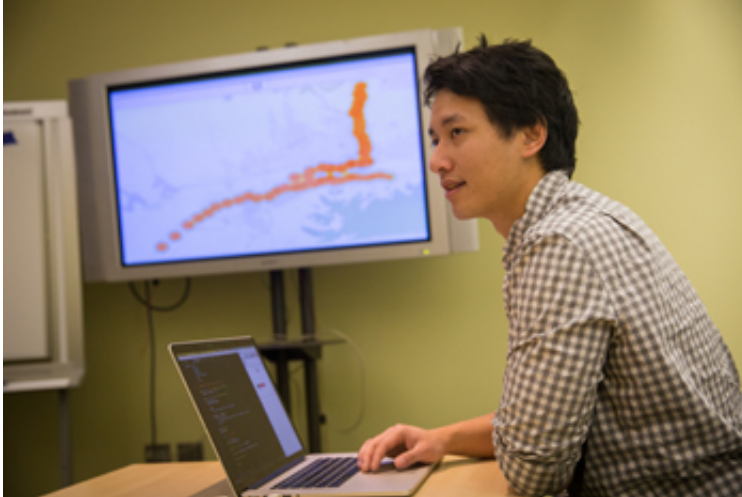


# Crowdsourcing a living map of world health

January 22 2014

---



CSE graduate student Andrew Huynh explains how data crowdsourced from "citizen sensors" around the world can be used to create maps of environmental pollution or infectious disease. Credit: Josh Knoff/UC San Diego Jacobs School of Engineering.

What if by collecting data from mobile medical apps on cell phones around the world, we could map significant problems and see the flu coming like a giant whirling hurricane? A team of engineers, biologists and medical researchers at the University of California, San Diego wants to leverage the widespread use of smart phone technology and cloud computing to build maps of large-scale health problems or environmental damage such as the concentration of heavy metals in drinking water. The idea is based on the principle that health, including infectious disease and environmental pollution, is a trackable geospatial

event. The team is working towards developing a tricorder that could monitor both individual and environmental health. In phase one, citizen sensors will test their drinking water using a simple test strip device that automatically sends the test results to a central data server for analysis while telling the tester whether the water is safe to drink.

They also hope their crowdsourced research will inspire support from individual investors through Indiegogo, the first partnership between UC San Diego and a crowdfunding platform. Examples include paying for the deployment of a sensor to an individual in one of five developing countries or buying one to monitor their own water plus that of an individual in another country. The project is led by Dr. Eliah Aronoff-Spencer of UC San Diego School of Medicine and Qualcomm Institute research scientist Albert Yu-Min Lin.

Their quest to create a living map of world health with the sensors, including infectious disease and [environmental pollution](#), was featured in a Jan. 14 story by Fast Company. We sat down with Andrew Huynh, a Ph.D. student in the Department of Computer Science and Engineering who is leading development of the team's data storage and analysis platform, to discuss the role of [machine learning](#) in the project. In machine learning, computers learn how to do something – how to distinguish a burial site from a pile of rocks in satellite images, for example – from the continuous input and analysis of data. When humans are in the loop as with crowdsourcing, Huynh says the picture can become hazy with inaccurate information. Part of his job is to train the computer to learn to distinguish accurate data from erroneous data.

## **Q: What is your role as "lead data scientist" on this project?**

As lead data scientist I helped architect and develop our open health

stack, a cloud analytics and storage platform that will be used to detect large-scale trends in data from sensors, individuals, and the environment. I also manage a small team of undergraduate researchers who work on various components as a way to gain knowledge of the latest best practices in programming.

**Q: How did you end up in this line of research?**



A "tricorder" developed by a team of researchers from Qualcomm Institute and UC San Diego School of Medicine features a simple test strip to check drinking water for pollution and an electronic device to check individual vital signs. The results are automatically collected in the cloud and analyzed to map environmental pollution or public health concerns. Credit: Alex Matthews/Qualcomm Institute.

I went into my Ph.D. program as part of the Valley of the Khans project, which was supported by the National Geographic Society and aimed at finding the tomb of Genghis Khan in Mongolia. The project involved a nondestructive archaeological survey utilizing modern digital tools from a variety of disciplines, including digital imagery, computer vision, nondestructive surveying, and on-site digital archaeology. I focused on

combining the application of machine learning to satellite images and human computation as a way to use [human perception](#) to help solve difficult problems.

**Q: You raise an interesting idea there. Using human perception to solve difficult problems? Do you mean that human perception adds something digital images and computers miss?**

Human perception is based on years upon years of complex pattern recognition and learning that we continue to do throughout our lives. There are many things are still very difficult for computers but very easy for humans, such as recognizing objects in images or understanding the intent behind a block of text. The concept of harnessing human perception, "human-based computation", uses the strengths of both computers and humans to achieve a symbiotic human-computer relationship to solve difficult problems.

**Q: Why is crowdsourcing research a big topic these days? What are the challenges in considering the reliability of the data?**

Researchers and businesses are beginning to see how crowdsourcing is a powerful tool to solve problems or obtain services, ideas, and content that would normally be completed by a traditional employee. The idea being that a collective of people, whether experts or non-experts, will give an aggregated result that is equal to or better than that of an individual expert. However, because these results are input by people who are prone to accidents or misperception rather than a deterministic machine, the quality of the data can vary from problem to problem often with no quantitative verification process. Determining the signal from

the noise in these huge datasets is a massive problem.

**Q: Is that your job as lead data scientist on this project to overcome this problem with the data? You're essentially teaching the computer how to distinguish good data from bad data, right? How do you do that?**

Absolutely. Once our sensors are deployed we'll be getting data from all over the world allowing us to experiment with different methods to distinguish good data from bad data when it comes from hundreds if not thousands of different sources. We're starting small first and then slowly ramping up to efficiently deal with each problem as we see it.

**Q: This shift towards a world in which individuals are sensors to be tracked and mapped raises considerable privacy questions. How do you think about privacy as a computer scientist trying to amass enough data to make inferences about trends in health and environmental pollution?**

Privacy is a concern in every aspect of this field, especially when you're talking about someone's individual health data. How do we anonymize this data so we can't even trace it back to the country of origin and yet still get enough information to be statistically significant? Also, in machine learning, we need raw data straight from the original source to train the machine. The process of anonymizing data removes subtle differences that are important in the machine learning process.

Anonymizing may tweak specific numbers or rip out whole sections of important information. Illnesses, for example, are often geographically

anchored. If we rip that information out, how do we learn anything? So it's an ongoing problem.

Provided by University of California - San Diego

Citation: Crowdsourcing a living map of world health (2014, January 22) retrieved 14 May 2024 from <https://phys.org/news/2014-01-crowdsourcing-world-health.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.