

VIEWPOINT

# Why crowdsourcing is not just for Wikipedia

BY PAUL LAW, CHERYL COHEN

22 JANUARY 2013

People power: The Interactive Autism Network collects self-report data from thousands of individuals with autism and their families, and provides a forum in which to discuss the results.

If you've seen a photo of a missing child on a milk carton, a Most Wanted poster or an article in Wikipedia, you've experienced crowdsourcing. Crowdsourcing is a way of distributing tasks to a large number of anonymous volunteers. It's an efficient method of accomplishing immense tasks with comparatively little investment.

With the advent of the Internet, crowdsourcing has become a common way of accomplishing large collaborative tasks such as **digitizing paper collections (reCaptcha)**, creating and maintaining encyclopedia entries, mapping and **cataloging the stars**, identifying and **tagging large numbers of images** or documents and **tracing the paths of neurons in the brain**. And yes, there is even a **crowdsourced list of crowdsourced projects**.

Crowdsourcing allows volunteers to become citizen scientists, archivists and journalists. With the Interactive Autism Network, we are applying the same power to autism research.

Crowdsourcing methods have already infiltrated healthcare and healthcare research. Among the most prominent crowdsourcing efforts in healthcare are **PatientsLikeMe** and **23andMe**. Individuals with rare disorders (more than 170,000 individuals reported) participate in PatientsLikeMe to track their symptoms and treatments and discuss issues they face. At 23andMe, people pay to have their

genomes analyzed and to receive reports about their ancestry and their risk for certain genetic disorders.

Each of these personal genetics companies has a participatory community component. During the registration process, 23andMe asks individuals for their consent to have their information and genetic data used for research purposes. A search of PubMed reveals that 22 journal publications have resulted from 23andMe data. A 2010 study, for example, showed that three different direct-to-consumer services provided highly similar data on DNA variants for one individual. However, they reported different disease risks to the consumer<sup>1</sup>.

## Autism network:

Our project, the **Interactive Autism Network** (IAN), provides a crowdsourcing mechanism for autism research. IAN has two components: IAN Research and IAN Community.

**IAN Research** is a scientific project in which individuals with autism and their families provide information in a secure online setting about items that are important to scientists, advocates and the community itself. To date, more than 43,000 individuals have joined the project. The registry collects parent- and self-report data consisting of baseline questionnaires, standardized instruments and topic-based surveys.

Although there is some concern about the validity of parent-report data collected via the Internet, IAN is able to reach many more individuals than the standard clinical model. The data enable IAN to contribute actively to scientific knowledge in autism, and have accelerated clinical research in autism by both IAN researchers and others.

Publications by IAN investigators alone include a population genetics study<sup>2</sup>, a study on regression<sup>3</sup>, a twin study<sup>4</sup> and a study on the stability of the autism diagnosis<sup>5</sup>. The committee for the upcoming *Diagnostic and Statistical Manual of Mental Disorders-5* is using three IAN studies to help refine diagnostic criteria for autism<sup>6,7</sup>.

The research is under the purview of the Johns Hopkins Medicine Institutional Review Board, so the participants have the same protection as those in other research projects. IAN shares information provided by participants in a **de-identified format** with qualified researchers throughout the world and through the National Institute of Mental Health's National Database for Autism Research.

All data are housed on secure servers at the institute and are not available via the Freedom of Information Act because the National Institutes of Health has issued a Certificate of Confidentiality.

IAN Research also matches willing individuals and families with appropriate local and national research projects. We inform participants when there is a study for which they qualify and provide

them with the researcher's contact information.

This provides the participant with the option of contacting the researcher and maintains the participant's privacy. In June 2011, we initiated a pay-for-services model in which we charge researchers a fee to cover our expenses. So far, we have helped recruit participants for more than 450 studies.

## New opportunities:

At the beginning of the project, researchers objected to our methodology because they did not believe that parent-reported information gathered over the Internet would be valid. To address this, we undertook several validation studies.

One, published in 2012, showed that there is a high degree of accuracy in parent-report autism diagnosis of verbal children when compared with gold-standard in-clinic diagnoses<sup>8</sup>. Another study found that although the families participating in IAN are not representative of the population, IAN has brought research opportunities to many U.S. families with autism who would not have otherwise participated<sup>9</sup>.

**IAN Community** is a website on which experts in public communication and particular topics provide information to the participants about the studies in which they have participated, and encourages the community to discuss the findings. This builds a feedback loop that allows participants to interact with the data via a set of data visualization tools. Using IAN **StateStats**, for example, a visitor can generate a graph from live IAN data that compares mean treatment costs in a selected state with the mean costs in the U.S. as a whole.

The aim is to give information back to the public so that they can learn about autism research, engage in it, join IAN Research and contribute to the research agenda. By giving current and potential research participants feedback about research, we hope to build social capital and trust. IAN Community has more than 700 articles.

We may be pushing the definition of crowdsourcing to include a **large-scale collaborative effort** involving many autism researchers, but it is important to also note the efforts of the National Database for Autism Research (NDAR).

NDAR consists of a technology platform, data standards and definitions, uniform identifiers and an ever-expanding aggregation of heterogeneous but linkable data contributed by many researchers and collected from more than 25,000 research participants (including more than 9,000 from IAN). NDAR allows researchers to access this massive dataset to test new hypotheses<sup>10</sup>.

Though this is different from crowdsourcing efforts to catalog the stars or digitize the world's libraries, this large-scale collaborative effort may yet result in important new discoveries.

*Paul Law is director of medical informatics at the Kennedy Krieger Institute in Baltimore, Maryland. Cheryl Cohen is director of online communities at the Kennedy Krieger Institute.*

## References:

- 1: Imai K. *et al. Clin. Chem.* **57**, 518-521 (2011) [PubMed](#)
- 2: Zhao X. *et al. Proc. Natl. Acad. Sci. USA* **104**, 12831-12836 (2007) [PubMed](#)
- 3: Kalb L.G. *et al. J. Autism Dev. Disord.* **40**,1389–1402 (2010) [PubMed](#)
- 4: Rosenberg R.E. *et al. Arch. Pediatr. Adolesc. Med.* **163**, 907–914 (2009) [PubMed](#)
- 5: Daniels A.M. *et al. J. Autism Dev. Disord.* **41**, 110–121 (2011) [PubMed](#)
- 6: Frazier T.W. *et al. J. Am. Acad. Child Adolesc. Psychiatry* **51**,28-40 (2012) [PubMed](#)
- 7: Rosenberg R.E. *et al.* **39**, 1099–1111 (2009) [PubMed](#)
- 8: Lee H. *et al.* **153B**,1119-1126 (2010) [PubMed](#)
- 9: Marvin A.R. *et al.* (2009, May). *Creating the digital melting pot: Lessons from a web-based national autism registry and research project.* Poster session presented at the International Meeting for Autism Research, Chicago, IL.
- 10: Hall D. *et al. Neuroinformatics* **10**, 331-339 (2012) [PubMed](#)