Meeting Summary

Bring Your Own Data: Opportunities and Challenges in Using Citizen-Generated Data

November 14th, 2012, 8:30am – 5:00pm
The Sheraton Crystal City

Hosted by:
U.S. Department of Defense
U.S. Department of Health and Human Services
U.S. Department of State
National Center for Food Protection and Defense
Alfred P. Sloan Foundation

The internet age has changed the landscape of how people communicate and share information. The ease of documenting events and changes in communities, neighborhoods, backyards, and homes through web-based applications and tools has improved public participation in scientific research and other projects. Citizen participation may include taking measurements, recording observations, and performing computations. The involvement of citizens can dramatically increase the scope, influence and impact of research projects.

The Bring Your Own Data Forum (BYOD Forum) brought together participants across several different subject areas (genomics, public health, ecology, seismology, and intelligence) to discuss different approaches and methods for using citizen-generated data and citizen science. The day proved informative and began to provide insight into some key questions and concepts.

Experts from the federal government considered how citizen-generated data could provide additional ways to accomplish their missions, and so partnered with the National Center for Food Protection and Defense and the Alfred P. Sloan Foundation to host a forum where experts could talk about public participation in data generation. The purpose of the BYOD Forum was to learn more about applications of citizen-generated data and how this data can help us understand what is going on in a community. For example, research projects or situational awareness efforts might benefit from a citizen science approach in areas where citizens are interested in the topic, are already collecting the data, or are looking to contribute to larger community/county/state/national projects.

Expert speakers were invited to provide insight and lessons learned. Their presentations addressed questions such as: Why are citizens participating in these projects? What is the quality of the data that is reported? And what challenges or limitations exist when using citizen-generated data? Highlights from each of their presentations were captured by note takers and are listed below. Some ideas are underlined which denote key elements, approaches, and lessons learned during the forum.

Welcoming Remarks:
Mr. Greg Delawie, Deputy Assistant Secretary, Bureau of Arms Control, Verification and Compliance, U.S. Department of State

Keynote:
Mr. Robert Ames, Senior Vice President, Information and Communications Technology, In-Q-Tel
http://www.iqt.org/

The extraordinarily large amounts of data now available have drawn the commercial sector headlong into the business of intelligence, with companies looking for highly directed ways to pique consumer interest. For data there are four V's: volume (amount of data), velocity (speed at which data is analyzed), variety (types of data – structured and unstructured) and veracity (trusting the data). Data today is massively unstructured and to analyze it, we need to change the way we collect and categorize it. The question remains on how we create innovation in data without breaking policy and privacy issues. For this data driven world, there is no single path or silver bullet to analytics and use, rather it is picking the right things to help solve the data gaps and challenges.

Transforming Research with Passive Citizen Data:

Dr. Jesse Lawrence, Stanford University (The Quake-Catcher Network)
http://qcn.stanford.edu/

The Quake Catcher Network (QCN) is a social-cyber-seismic network made possible by volunteer computing. The goal is to increase the number of seismic sensors by using volunteered distributed computing with low-cost microelectromechanical accelerometers. Collaboration is important as demonstrated by the QCN merging the cyber infrastructure, seismology, education, and engineering communities for improved results. The QCN is looking to bring in more participation from schools and classrooms, and they provide volunteers with information in exchange for participation.

Dr. William McShea, Smithsonian National Zoo
http://siwild.si.edu/

The Smithsonian Wild Camera Trap project uses motion triggered camera traps to photograph animals to further conservation and ecological understanding. The public can participate in this project by helping to set up these camera traps and submitting photographs to the Smithsonian Wild website. Volunteers attend a training session on the project, receive an assigned area to cover, upload a desktop app for the traps, send photos for expert review and image repository, and then can have their photos posted to the website. The public participation for the project is a tiered volunteer system with some personal contact from project coordinators and public participants.

Bring Your Own Data – Participatory Data Collection:

Dr. Rumi Chunara, HealthMap (Flu Near You)
https://flunearyou.org/

Flu Near You is free website with available apps that aggregates information on influenza. Volunteers sign up to be sentinels for influenza and receive email updates on the flu information in their areas. Flu Near You keeps participants engaged through surveys sent out every Monday morning that ask simple questions, requiring only a few minutes to complete. The project aims to provide a visual platform to engage people worldwide on disease surveillance.

Dr. Julie Urban, Nature Research Center, N.C. Museum of Natural Sciences (Your Wild Life)
www.yourwildlife.org

Your Wild Life reaches out to the public to help conduct research on skin associated bacteria. Your Wild Life takes skin samples from the belly buttons and armpits of museum visitors to help understand why bacteria differ from person to person. The project involves the public by sharing of data and by making the research process visible through a glass window into the research lab at the museum. The challenges this project has faced include the length of time to publish the findings, needing to relate findings in a meaningful way, and personal privacy issues.

Ms. Ruth Breech, Global Community Monitor (The Bucket Brigade)
www.gcmonitor.org

Global Community Monitor's project, the Bucket Brigade, provides air quality monitoring devices and training to local volunteers who test gases and particulates in the air. Community members are encouraged to establish a trail of evidence by recording what they see, smell, and feel in the air in their communities. Local adaption of the Bucket Brigade provides a way for the

community to have ownership of the project and encourages the community members to incorporate innovative ideas. Partnerships as well as capacity building have been critical for the project.

Building Situational Awareness With Passive and Participatory Data:

Ms. Megan Hines, University of Wisconsin Madison School of Veterinary Medicine
Wildlife Data Integration Network
http://www.wdin.org/
Wildlife Health Event Reporter
www.wher.org

The goal of the Wildlife Data Integration Network (WDIN) project is to locate, organize and standardize information to present to decision makers on the topic of wildlife health. The WDIN utilizes different techniques to locate and filter a variety of sources including news reports, citizen derived information, official information and social media posts. A specific example of their efforts is the Wildlife Health Event Reporter (WHER), which is a web-based app that captures observations about dead and sick wildlife, provides near real-time alerting, and can be used as a destination for reports by anyone, from the general public to resource managers and other health agencies. The WDIN has discovered that many challenges to wildlife data integration exist in addition to typical data integration challenges: there is no mandate to share information; there are few accepted and adopted standards for data; there is a disconnect between information technology and information management; and there are bureaucratic, research and funding challenges.

Cross-Disciplinary Dialogues for Sharing and Collaborating:

Ms. Jennifer Shirk, Project Leader, CitizenScience.org, Cornell Lab of Ornithology
www.citizenscience.org

The field of citizen science is rapidly growing. Innovation, networking, and advocacy of the field, as a whole, must occur to build a professional field of citizen science best practices. CitizenScience.org provides a place for those interested in the field to find resources and to register their projects to see what kinds of activities are possible. Citizen science projects should be as easy as possible and show the public the value added from their participation. Projects should encourage people to share information on their work and to publish.

ASPR Challenge Competition Winner Announcement:
RADM Nicole Lurie, MD
MappyHealth team (Charles Boicey, Brian Norris, and Mark Silverberg)
http://mappyhealth.com/

HHS/ASPR's Now Trending Challenge winners, MappyHealth, used the Challenge hashtag #nowtrending2012 to connect on Twitter, leading to a collaboration that built the winning app for tracking the top five health issues trending in an area. The winners have taken their collaboration and formed a new company called Social Health Insights and are planning to expand their platform's capability. The Mappyhealth tool shows that unstructured public social media data can be captured and visualized for greater understanding of disease trends and topics.

What can we learn from Citizens generating ecological data:

Dr. Sandra Henderson, Project Budburst
http://neoninc.org/budburst/

Project Budburst looks at plant life cycles and tracks the impact of climate change on plants. Their primary goal is education and outreach to a broad audience on their issues with a secondary goal of obtaining useful data. They provide online courses and education resources for volunteers and the public. They have found partnerships to be key to success, and their identified challenges include recruitment and retention of volunteers.

Mr. Timothy Dye, Sonoma Technology Inc.
http://www.sonomatech.com/
http://airnow.gov/

Sonoma Technology supports the Environmental Protection Agency's AirNow initiative looking at air quality in different areas. Recently, AirNow looked at expanding their coverage of air quality monitors to local high schools through the Global Ozone Program (GO3), which does similar air quality monitoring work. They discovered that inclusion of the GO3 data reduced errors in their own monitoring capability, increased area of coverage and served a larger population area. Another project Sonoma Technology supported was Street Sensing (citizens taking measurements with small air monitors), which was part of an urban prototyping contest to bolster civic participation to improve city life.

Dr. Anna Switzer, National Geographic Society, National Geographic FieldScope
http://education.nationalgeographic.com/education/program/fieldscope/?ar_a=1

FieldScope is a geographic information system - and web-based data mapping and analysis platform to help citizen scientists share and explore the data they have collected. Soon, but not yet, the platform will provide project builder tools to help individuals and organizations do their own investigations - locally and nationally. The project encourages collaborative data sharing and analysis within the user community. One example project is the Yukon River Inter-Tribal Watershed Council that works to monitor the health of the Yukon River including removal of trash and testing for pollutants.

What can we learn from Citizens sharing genomic and personal data?

Mr. Greg Biggers, Genomera
http://genomera.com/

Genomera focuses on networked health studies by empowering participants to ask and answer research questions with participation from others in the network. They support participatory driven clinical trials and health studies. Their work looks at online study operations including study design, protocol tracking and data collection, and participant engagement including recruitment, enrollment, engagement and re-enrollment. By conducting participatory driven clinical trials, participants feel that the locus of control of the study is with them and that they see much more transparency in the results of the study.

Mr. Jason Bobe, PersonalGenomes.org
http://www.personalgenomes.org/

The Personal Genomes Project (PGP) studies the combination of genomics and environment to discover traits. The PGP is creating a different kind of human health and disease model by building a cohort of individuals who are a safe test bed (people who consent to public access of their health data) to become the most analyzed humans on Earth. PGP platforms encourage sharing data and information, from participants sharing the results to researchers distributing information to the participants. The project is looking to expand beyond the United States and seeks investigators who will study a particular trait in an individual to research.

Dr. Jérôme Waldispühl, McGill University (Phylo)
http://phylo.cs.mcgill.ca/

Phylo is a game developed to help decipher genomes across various species. For Phylo, using gaming technology has allowed researchers to improve analysis of genomic regions. Phylo plans to expand its players to other countries and to mobile device platforms. The key is to bridge scientists and researchers to gamers and game developers to find the right subject, model or problem that would lend itself to being solved or assisted by a game.

How Federal Agencies and Federal Partners are Using Citizen-Generated Data:

Ms. Nancy Nurthen, Department of Defense (DTRA)
Dr. Jason Matheny, Intelligence Advanced Research Projects Activity (IARPA)
Dr. Peter Preuss, U.S. Environmental Protection Agency (EPA)
Dr. Amy Kircher, National Center for Food Protection and Defense (NCFPD)

Highlights from the Federal Agency panel:

- DTRA is looking to build an ecosystem that looks at biosurveillance data, collaboration, analytics and visualization. (http://www.defense.gov/news/newsarticle.aspx?id=117673)
- IARPA is looking at a range of ideas to develop open source indicators to monitor protests, riots, labor strikes, elections, disease outbreaks, and economic issues. For example changes in Wikipedia articles may indicate a changing event. (http://www.iarpa.gov/Programs/ia/OSI/osi.html#)
- EPA has re-discovered citizen science. They are hosting a competition with HHS to look at sensors and apps. They are also looking to help communities around refineries and chemical plants on ways to monitor benzene and other chemicals in the air. They hope to expand such efforts to a range of chemicals and to monitoring pollutants in the water. While citizen science is a good place to begin outreach, EPA is also interested in the next steps of connectivity with "sensornets." (http://epa.gov/research/challenges/)
- NCFPD is a federally-funded Center of Excellence that is tasked with protecting the food system and keeping it safe from catastrophes (adulteration, terrorism, economic). They are looking to use citizen-generated data to help shift the epidemic curve to the left so events are caught before people start getting sick from the food they have eaten and require healthcare. (http://www.ncfpd.umn.edu/)

Conclusions

Citizen participation can bring great value to projects by increasing the number of data points collected, serving a larger population area, and promoting public awareness and concern for specific issues. Citizens are more likely to participate in studies focusing on a subject matter that they are already passionate about. Citizens want to know that someone is listening to them, helping them feel engaged and empowered. Projects need to be open and transparent with the locus of control with the participants. Citizens want to see a visible research process with intermediate results shared along the way and final results shared in ways that are understandable. Participation in the project should be easy, with some added value seen in projects that incorporate an online community and educational materials. At the end of it all, citizens want to get something back from the project such as further education, visualizations of their data, or participation in future studies.

There are certain challenges and limitations to citizen-generated data that were discussed during the forum. Issues with vetting and introduction of bias into the data can exist as researchers and project coordinators make decisions on what data to accept/reject and how to correlate the data with the number and location of participants. Recruitment and retention of citizen volunteers can require significant time and planning for the project team. Further, unique privacy and legal issues may exist with citizen-generated data which must be addressed before the data can be collected.

As this data source is further explored by the participating federal agencies, information gathered in this forum will help guide the next steps for integrating this data source into the work we do. Advantages and limitations explored in the forum will help the participating federal agencies craft our approaches and strategies. Further, connections made with the expert speakers will open the door to continued conversation and dialogue on citizen participation in data collection.