Data, Data, Data

Currently, most environmental health data are siloed in local departments. Some of it might get shared with our state programs in some instances. We store information in Access, HealthSpace, Accela, Custom Data Processing, Inspect2GO, RedCap, and our alphabet soup of databases. I would like you to see these databases as cell phone towers. The more towers in a network, the better the coverage. These databases can have the same impact if they are all networked together. As new data are added to our network of environmental health data, the gaps become smaller. As a disease or pollutant crosses the country, our data would be there to help us more quickly recognize and fix a problem. Currently, it is difficult to find data and slower to retrieve it if it is kept on a different network in a different county or state. If only we had a way to build an information network that connects with adjoining counties, states, and even the federal government.

Public health has proven over the last couple of years that the technology is available to make this type of network happen. For example, consider the >500 million COVID-19 vaccines administered and the millions of COVID-19 test results recorded in a little over 1.5 years. All these data have been captured locally and shared with state agencies and CDC. From there, the information is transferred to universities such as Johns Hopkins and reported in The New York Times overnight. If we can achieve this sharing with COVID-19 data, we can do it with other data.

Another example of a public health network is the Community Well-Being Index, a website that reports on nearly 600 data points broken down by almost every county in the U.S. (https://wellbeingindex.sharecare.com). The information is then shared openly around the country to improve health. Environmental health can and should learn from these examples to build a vast network of data that will help create a more informed workforce for our communities.

I am not suggesting that everyone use the same database or put all their data on an open-facing database. In many cases, we are not able to do that for legal reasons. We must, however, start looking at the obstacles standing in the way of quick and easy data sharing capabilities. The first thing that needs to change is our mindset that data are ours and we must protect it at all costs. That is crazy and violates our social philosophy to do the most good for the most significant number of people. There are many lines on a map and we see that our world is getting smaller each day. Lines on a map do not stop the progression of disease and illness, but it has built barriers in our minds and policies for decades. Foodborne illnesses spread rapidly across the country. West Nile virus spread from coast to coast in only a couple of years. To combat what ails us, we need to know more.

I am not an information technology specialist but having worked with the Partnership for Food Protection to create an integrated food safety system, I have learned that to share data, creating a data dictionary...
is crucial. The data dictionary builds a foundation so that different databases speak the same language. FDA has started doing some of this work in the manufactured foods arena.

Once information is given a common name, it must be stored so that others can access the data. Access to data needs to be down two-ways or in a push–pull manner. Users must provide and retrieve the information, or they will not feel that it is of any use to participate in sharing. Formal agreements and metadata can be created to communicate what participants or data users are getting and why the information was created. Policies, rules, and regulations can prevent the misuse of these data.

While we learn to share data, we also need to teach our workforce how to use the power of this information. Many environmental health professionals are great at collecting information, yet most do not analyze and interpret what they have. Collaboration with research facilities and academia can teach us how to use data to make better predictions on water and air contamination, where we are most likely to have a foodborne illness, and when and where the next disease outbreak will occur. Information can even be used to determine workforce shortages or how to best use existing resources to prevent an outbreak.

Having good data will also help environmental health professionals gain a seat at the policy making table. Healthcare professionals, community planners, elected officials, and many others need data when making decisions. If we have data they seek, we will be invited to share it. Many of us already experience this inclusion in the process of public health department accreditation as environmental health data are used to help meet several accreditation requirements.

As I write this column, many of us in the NEHA leadership and staff have discussed creating some type of data lake. What information can we capture? How and where can we store it? Who will benefit from the information? How will the data improve our profession and the health of the public? Can those professions and individuals outside of public health benefit? The possibilities are endless. I say that if you build the network, the users will come.

Let me know what you think. Is this too futuristic? Is the thought of data collection, storage, and use essential, or is it a waste of time and money?