‣ **BUILDING CAPACITY**

Downstream Data Flows to Build Capacity

Darryl Booth, MBA

Editor's Note: **A need exists within environmental health agencies to increase their capacity to perform in an environment of diminishing resources. With limited resources and increasing demands, we need to seek new approaches to the business of environmental health. Acutely aware of these challenges, NEHA has initiated a partnership with Accela called Building Capacity—a joint effort to educate, reinforce, and build upon successes within the profession using technology to improve efficiency and extend the impact of environmental health agencies.**

The *Journal* is pleased to publish this column from Accela that will provide readers with insight into the Building Capacity initiative, as well as be a conduit for fostering the capacity building of environmental health agencies across the country. The conclusions of this column are those of the author(s) and do not necessarily represent the views of NEHA.

Darryl Booth is the general manager of environmental health at Accela and has been monitoring regulatory and data tracking needs of agencies across the U.S. for almost 20 years. He serves as technical advisor to NEHA's informatics and technology section.

Many health departments publish their routine restaurant inspections to the Internet. On a simple web form, consumers can easily search by a restaurant's name to find inspection history, inspector comments, violations or inspection checklists, and details on an establishment's score, grade, or placard (where applicable). A practice once considered sensitive—and possibly even controversial—is now commonplace.

There are some unexpected ways, however, in which these data can be repackaged.

Open Data
A few health departments have embraced the open data philosophy by publishing inspections and other data files online for bulk download or quick consumption through digital tools like application programming interfaces (APIs). This open data publishing practice—an ideology that embraces the idea that data should be freely used, reused and redistributed by anyone—allows any software, person, or company to ingest and repackage the data, presumably adding value to them through the process.

Yelp advocates for this practice, having published a standard file format known as Local Inspector Value-Entry Specification (LIVES). Health departments that put their inspection data in this prescribed format could expect their inspection scores or grades to be visible alongside Yelp reviews.

In a searchable website intended for consumers, health departments can create and dictate the brand, presentation, and search method of their data. A health department can inject its philosophy, framing the results with useful articles about safe operations and consumer awareness. Let's call this site a "Web Search."

With an open data API or bulk download, data are repackaged with little to no expectation of a health department's specific perspective. In fact, if the data are combined with disparate data from other sources (e.g., other health departments and consumer reviews), the data's origin is clouded. I'll call this practice "Bulk Download."

Screen Scraping
Did you know that your Web Search site is likely being turned into a Bulk Download? It happens through screen scraping. Screen scraping is a technique through which computer software navigates your website by pretending to be a human clicking links. Each time your website serves up a snippet of data (i.e., a web page), the screen scraping program grabs that text and adds it to a database. The process might take many hours or days and generate hundreds of thousands of simulated clicks to your website. When every click is exhausted, all the data on your Web Search site is consolidated into a massive database to be sold or repackaged.

Screen scraping can be a tricky proposition and is generally not preferred because the program has to be aligned to your website's design. That is, it has to know where to find the facility name, facility address, violations, and score (if applicable). It also has to know where to click to get the next page of results and so forth, until it reaches the last link on the last facility. A Bulk Download or API request is much more preferred.

Consolidated Data

I believe it's a problem that no national accounting of retail food safety inspection results exists. I also believe it's a problem that scoring/grading across the U.S. lacks a standard methodology. It's confusing to consumers and problematic for multiregional owners (i.e., business with facilities in many different jurisdictions).

Some firms seek to solve these problems by collecting data from different local jurisdictions. They then consolidate the data by calculating restaurant inspection scores through their own scoring methods.

The final destination for these scores? Yelp, for one, licenses a data stream from HDScores. HDScores scours a good portion of the U.S. and boasts 1.3 million restaurants in its database. If a local health department calculates a score, then the health department score is visible on Yelp. Otherwise, the proprietary HDScore is displayed in Yelp, alongside consumer reviews. The HDScores calculation weighs both current and historical inspection results. Details can be found on the HDScores website at https://hdscores.com.

Another example is Hazel Analytics (https://hazelanalytics.com). Hazel Analytics publishes food inspection scores through its Food Safety Insight, a food inspection analytics system that benefits corporate brands such as Subway, Starbucks, and Red Robin. The value proposition for these customers is homogenized results that are displayed in a dashboard of actionable alerts for their facilities.

Operational Matters

It is possible that, for good reason, a set of data should be excluded from consolidation. There are two methods to prevent the automatic scraping of your agency's data.
1. Identify and contact the companies doing the scraping/consolidation. With just a couple phone calls, I confirmed that the companies will honor most requests for exclusion.
2. Work with your information technology (IT) department or software vendor to block originating screen scraping requests from reaching your web server. This method is similar to blocking an unwanted caller on your mobile phone but is imperfect in the sense that the originating computer system could change in an endless game of cat-and-mouse.

If, however, your Web Search site suffers from an undo amount of traffic due to screen scraping (screen scraping can cause web traffic spikes as the process hits every link on your website), your IT department or software vendor can offer to publish a Bulk Download link or API, thus optimizing the regular transactions and reasserting your control over the data stream.

At the end of the day, publicly available data can still be requested through public records requests.

At a minimum, a health department should assure that flow (no matter what the method) is updated at least weekly and that operators have easy access to identify errors and that errors are corrected promptly.

Join the discussion on the LinkedIn Building Capacity in Environmental Health group at www.linkedin.com/groups/6945520/. 🕶

Corresponding Author: Darryl Booth, General Manager, Environmental Health, Accela, 2633 Camino Ramon #500, San Ramon, CA 94583. E-mail: dbooth@accela.com.

Did You Know?

NEHA has transitioned its online courses, trainings, webinars, and other resources to an upgraded learning management system. The new platform is free to all active NEHA members and can be used to earn continuing education credits toward a NEHA credential. The platform provides an improved interface to find and access resources right away. Visit www.neha.org/professional-development/education-and-training/neha-e-learning to learn how to access the new E-Learning platform and to take advantage of these resources.

Find a Job | Fill a Job

Where the "best of the best" consult...

NEHA's Career Center

First job listing **FREE** for city, county, and state health departments with a NEHA member and for active NEHA educational and sustaining members.

For more information, please visit neha.org/careers.