Attn: Rep Bryan Cutler FYI – First Draft Thursday, December 08, 2011

This Data Scrubbing article was sent to EPA HQ in Washington, DC for their review, in the hope of getting a co-authorship, by a federal agency before publication.

Article Word Count 1,877

Scrubbing Source Data at the Local Level

By Michael Jacoby

First responders and local private citizens are the first line of defense in times of crisis or need. Ensuring that those individuals and response units are equipped with accurate and reliable information is imperative. Computer data errors and/or discrepancies in name, address, site locational data, contact information, phone numbers, etc. during an event can lead to a response unit being dispatched to the wrong location, or responders and local citizens being unaware of hazardous conditions that require special attention in their local communities. Waiting for out-of-area assistance to arrive, or for the initial responders to be re-routed to the correct location, could mean the loss not only of valuable minutes but also, in some situations, of human lives.

The Envirofacts website of the Environmental Protection Agency, EPA says, "provides access to several EPA databases to provide you with information about environmental activities that may affect air, water, and land anywhere in the United States [emphasis added]." Those "activities" include but are not necessarily limited to "toxic chemical releases, water discharge permit compliance, hazardous waste handling processes, Superfund status, and air emission estimates." Over the past few years, though, numerous locational errors have been found in the database itself, and most all responders and planners referred to this type of locational data/information as being Facility Registry System (FRS) information. After the provided addresses were plotted by the EPA based on information collected or provided data, a disturbing large number of various "sites of interest" in certain areas were found to be positioned in the middle of intersections, on interstate highways, and even in farm fields. Among the other erroneous data found were a number of properties some plotted as much as 20 to 40 miles or more away from their correct locations.

These data discrepancies were brought to the attention of U.S. Representative Todd R. Platts (R-Pa.), EPA Administrator Lisa P. Jackson, the Department of Homeland Security (DHS) FEMA

United States Fire Administration (USFA), those involved with Centers for Disease Control and Prevention (CDC) discussion groups, as well as members of various private sector groups, organizations etc.. On 7 January 2011, EPA Assistant Administrator/Chief Information Officer Malcolm Jackson. shared my view "that the EPA data is vital for the public, and should be as accurate as possible."

It is usually assumed, of course, that official databases such as Envirofacts are indeed "factual." The problem with that assumption is that the data provided can only be as accurate as the data entered, and the locational source data for certain sites of interest is provided by a broad spectrum of state and local government entities, stakeholders etc. such as state departments of labor, environmental departments that may have different filing and data requirements. It is for that reason, that it is imperative that the information being provided by government systems, and shared not only with emergency services agencies but also with the general public, be as accurate as possible through the incorporation of a rigorous validation process. However, verifying and updating this vital data requires a much more, and more effective, public-private collaboration – on a continuing basis in order to address the deficiencies within the system.

Millions of Records – Each and Every One is "Unique"

According to the EPA website, the FRS now has available "over 2.5 million unique facility records linking over 3.0 million program interests, including data from over 25 national environmental data systems and over 45 state systems." After numerous examples of locational errors had been brought to the attention of the EPA, my state and other government officials, and not until after seeing examples firsthand, did a data-scrubbing process start in south-central Pennsylvania with the goal to ensure that the locational data for "site of interest" in that area would be accurate and complete. The process that they decided to use was of their choosing.

Obviously, knowing how to check this vital locational data and how to Report an Error to the EPA could help reduce potential delays during future emergency-response efforts, if replica data is showing the same locational errors. When creating a risk management plan (RMP), therefore, it is particularly important to check the vital site locational information already available for local facilities to ensure that such shared data is both accurate and up-to-date. In addition to the dangers that can affect the general public, there are also many cases where exposure to a substance may affect only a select group of individuals not recognized by other entities because it's out of their domain of control. For example, persons with "special needs" – or suffering from

hypersensitivity or from allergic concerns to certain chemicals may need additional assistance if those same persons are living or working near one of the facilities listed / having an RMP.

Some local governments maintain lists of special needs residents – e.g., ECRIN is used in York County, Pa., to Evacuate County Residents In Need. Other persons, afflicted with an even higher level of sensitivity, might already be on a state's Hypersensitivity Registry list recognized by other departments etc. Having those lists available can help the response efforts considerably in sudden times of crisis. That is why it's imperative that the site FRS locational data is accurate and up-to-date, depending on the wind direction and speed during an event.

DV, OTIS, OSWER, VZIS.

The first step needed to correct current government data is to acquire basic knowledge about Data Verification (DV) procedures. A government employee sitting at his or her desk at EPA Headquarters in Washington, D.C., cannot, at present, accurately determine whether a site's locational data is correct – because that information usually can be verified only at the local level by persons familiar with the site's correct location. To rectify the errors discovered when incorrect (and/or incomplete) data is reviewed (and/or verified), the federal government has established a process, managed by the EPA, to Report an Error by using the EPA's Integrated Error Correction Process and Error Tracking System (ETS). Among the principal users of such data are the EPA's Office of Solid Waste and Emergency Response (OSWER) and other agencies and departments that draw information from this authoritative database system. Other users of this vital information are organizations such as Environmental Justice groups, and – the most important of all, the general public who have loved ones in the communities.

Another tool offered through the EPA website by the Office of Emergency Management is the Vulnerable Zone Indicator System (VZIS), which offers a quick way to determine if a particular location that might be affected by a chemical accident and/or is in the "vulnerable zone" of a facility submitting an RMP. The 1986 Emergency Planning and Community Right-to-Know Act and certain chemical accident "prevention provisions" in the 1990 Clean Air Act Amendment help ensure that certain information on possible hazardous chemicals stored/warehoused at certain businesses and/or other local facilities is publicly available from state and local governments. From a responder perspective you know this type of information from CAMEO. Please note this is not the detailed quantity specifications within the facility or site of interest that many first responders use, but rather the basic locational data, FRS data, directing you to the

proper location in times of crisis. If shared data within data systems has you going in the wrong direction everybody… has a problem!

True community preparedness requires the earnest and continuing efforts of all persons who live and work within the boundaries of that same community. When information related to various sites of interest in the community are in error – e.g., plotted in the wrong location, or with incomplete or incorrect contact information, including phone numbers and addresses, etc. – the EPA's reporting process can help significantly not only in reducing the reporting times required for individual citizens but also decreasing the processing time required and needed by the EPA to correct any errors within the system.

One example: After seeing the amount of locational errors in south-central Pennsylvania that were researched and verified, officials of York County committed to scrubbing the EPA data for their jurisdiction, as already listed- in alphabetical order. This process will hopefully be a one time scrubbing. By learning more about the reporting process as you look at the accuracy of federal database systems, other local governments, agencies, and private citizens can determine if the data for their own facilities and sites of interest should also be thoroughly scrubbed. Restoring the trust in data systems that are used in times of crisis or unusual need must be a whole-community effort if total community protection is the goal that must be attained.

To verify and correct information for sites of interest in FRS, use the following procedure: (a) visit the EPA – Envirofacts – Multisystem Query site this site is unique and has hyphens between the words http://www.epa.gov/envirofw/html/multisystem_query_java_bk.html; (b) in the "Geography Search" area, enter a local-area Zip Code number; Note: Start with a ***rural*** *ZIP code number area until you are familiar with the use and operation of the system (c) click "View Facility Information" next to the known facility name and address; (d) if the mapped location of the facility is incorrect, click the* ***"Report an Error"*** *button in the top right corner of the page; and (e) follow the instructions provided by the EPA. When the address does not match the facility address assigned, by using geo-spatial data GIS latitude and longitude coordinates locational information you can quickly solve these problems in seconds by using a siting tool or a basic GPS device. Contact me and I will show you the application.*

The Vulnerable Zone Indicator System (VZIS) is also available at: http://www.epa.gov/OEM/content/vzis.htm After you receive your initial report with your boots on the ground experience you will be able to quickly test the VZIS system by comparing their information to what is in your first due response area that is already on record. Don't think about yourself because you will have your PPE. Think about the other members of your community who use this data that have children or loved ones with special needs. It is extremely important to compare your information and then make the necessary corrections! Your neighbors and friends might already be on a state's Hypersensitivity Registry and unbeknownst to them the site locational data could be wrong. If discovered in error immediately bring the to the EPA's attention by Reporting an Error. If you discover numerous errors do screenshots of your findings and send them to your federal representatives for their review.

Article on "EPA Errors on Environmental Hazards Map Send York County Man – And Government – On a Quest" available at: http://www.ydr.com/ci_19121036

_____

Michael Jacoby, is a private citizen from York County, Pennsylvania, EPA Region III, who provides "awareness information" about data scrubbing to the public.

If we lose our first responders because of complacency and computer data errors within shared data systems the individuals that could be affected the most could be those within your local community.

For years many first responders did not know how to check the data within the system to protect their own. Now you have the knowledge to do so!

Note to DomPrep Editor: **Insert** the Flowchart from EPA's website to describe the Error Correction Process (http://www.epa.gov/enviro/html/error/flow_chart.htm) if graphic space is available.