GRADIENT

TRENDS

Risk Science & Application

Fall 2015
Issue 64

The Role of Big Data in Solving Environmental Problems

By Jeff Rominger, Ph.D. and Shingo Ikeda, M.A.

Big data can create big challenges for scientists related to the "four V's" – volume, variety, velocity, and veracity.

Over the last decade, an increasing number of environmental issues have required the analysis of data from hundreds or thousands of different locations, drawing on regional, national, and global monitoring networks. These comprehensive data sets can provide crucial context for comparing current to historical conditions at a specific site. For other environmental issues, such as atmospheric pollution, the "site" is an entire region, nation, or the world, requiring a holistic assessment using large scale data sets. As a result of the massive volumes and wide variety of data types available now to scientists, solutions increasingly require the tools and techniques of "big data."

The term big data is overused, often ill-defined, and creates confusion even among scientists.

The term big data is overused, often ill-defined, and creates confusion even among scientists. Although there is no strict definition, big data is often characterized by traits known as the "four V's" (IBM Corp. 2014): volume, variety, velocity, and veracity, which are defined as follows.

- Volume, or the size of the data set renders traditional data storage and computing tools obsolete.

continued on pg. 2

I N S I D E

Letter from the Editors

September 2015

Dear Colleague,

This issue of *Trends* focuses on geographic information systems (GIS) and environmental data management systems, which, over the last decade, have become integral to analyzing and solving complex environmental issues. These systems can be powerful decision-making tools because they allow users to consider and analyze reams of environmental data at near instantaneous rates. However, relying on these systems and using such large data sets present a host of new challenges.

The first article looks at the role so-called "big data" plays in solving various environmental issues and some of the challenges that can arise. Our second article discusses how GIS databases can be effective tools in environmental and toxic tort litigation. The third article discusses geospatial methods used in environmental epidemiology, including disease mapping, cluster analyses, and geographic correlation studies.

Gradient contributors to this issue include Dr. Jeff Rominger, Mr. Shingo Ikeda, M.A., as well as Mr. Matthew Mayo, M.S., GISP, CPG, P.G., Ms. Naomi Slagowski, M.S., and Dr. Christine Loftus, M.S., M.P.H. Joining us with a guest editorial on the use of GIS databases in litigation is Christopher Thatch, a commercial litigator at the law firm Jones Day.

We hope that this issue of *Trends* provides you with insight on this topic.

Yours truly,

Kurt Herman, M.Eng., P.G. and Chris Long, Sc.D., DABT
kherman@gradientcorp.com clong@gradientcorp.com

GRADIENT
Trends is a free publication of Gradient

The Role of Big Data in Solving Environmental Problems

continued from pg. 1

- Variety, meaning multiple types of data such as satellite images and maps, time series of chemical concentrations, wind speed, air quality, *etc.*, all of which must be appropriately linked together.
- Velocity, meaning data can now be accessed in near real-time from wireless monitoring networks across the country.
- Veracity, meaning that the size and complexity of the data requires new quality assurance/quality control techniques and algorithms to ensure data quality and accuracy.

A simple, illustrative example of the value of big data is a retailer's vast records of customer transactions, which can number in the millions. A retailer can use the patterns in their transaction records to develop new products, set prices, or deliver targeted advertisements to customers. As the data set grows larger, more potential patterns can be identified. Environmental data sets can be similar in scale, but are typically much more complex, often encompassing disparate types of data linked by complex physical or chemical processes. For example, evaluating the potential effect of human activity on groundwater quality throughout the United States would require gathering information on massive monitoring well networks (*e.g.*, depth, construction characteristics), multiple measurements of water levels and chemistry in each well, and maps and statistics on the variety of human activities throughout individual watersheds (see graphic). These data records would include maps, watershed boundaries, historical time series, well logs, *etc.*, and therefore the ability to compile, organize, and analyze such data sets would inevitably require specialized data management skills.

LOCATIONS OF GROUNDWATER LEVEL MEASUREMENTS IN THE NORTHEASTERN U.S.

Source: USGS, 2015. The USGS maintains a database of groundwater level measurements at more than 850,000 locations throughout the U.S. http://nwis.waterdata.usgs.gov/nwis/gwlevels.

As big data becomes more ubiquitous, it is essential to understand all of the challenges of working with big data in environmental issues. The volume of data presents the most immediate challenge as data sizes exceed the limits of most machines and software. To address this issue, data sets are increasingly stored in a distributed manner on cloud networks, which alleviates storage volume limitations and also allows remote data access. Additionally, the variety of data types and the need to link these data together (*e.g.*, chemical concentrations and locations in aerial photographs) in a logical way is an ongoing challenge for scientists. The velocity of data throughput in both incoming and outgoing streams is magnified with big data. Many types of data input are becoming *streaming* instead of being traditional *static* files, meaning that the data are produced and delivered in near real-time. Data processors are required not only to capture the data flow, but also to analyze data at the same time. This concept is not new and many data such as weather and oceanographic data are delivered in streaming format; however, new big data technologies can process these data streams at incredible speed and integrate the data into informative graphics, web pages, or simple figures, such that scientists and non-scientists alike can digest and use the incoming information. Finally, one of the biggest and most fundamental challenges with any data set is ensuring the veracity of the data. The adage of "garbage in, garbage out" is especially true for big data as the data sets are too large to be checked by hand, and automated algorithms must be used to identify duplicates, omissions, or errors.

Without reliable data, there can be little confidence in scientific conclusions. Given the volume and variety of environmental data available for analysis, and the speed at which these data are created and accessed, the need for big data expertise coupled with expertise in the environmental sciences will continue to grow.

The authors can be reached at jrominger@gradientcorp.com and sikeda@gradientcorp.com.

References:

IBM Corp. 2014. The Four V's of Big Data. Accessed on July 20, 2015 at http://www.ibmbigdatahub.com/infographic/four-vs-big-data.

GIS for Environmental Litigation

By Matthew Mayo, M.S., GISP, CPG, P.G., Naomi Slagowski, M.S.

A GIS database can be an efficient and cost effective tool for organizing and assessing the huge quantities of information generated in environmental litigation.

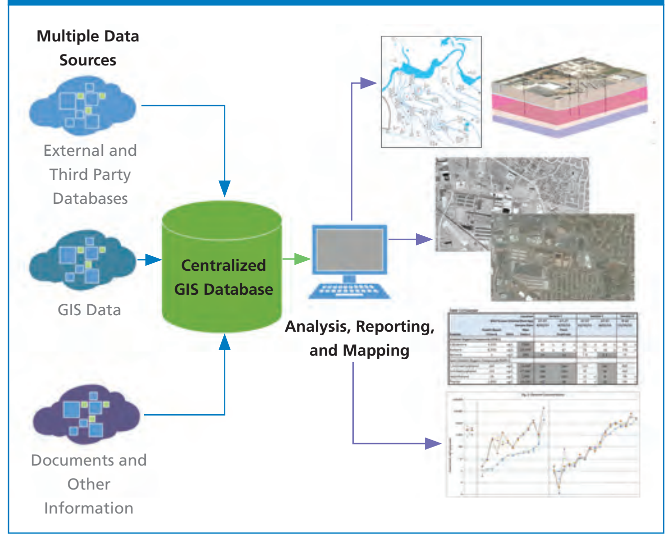
A Geographic Information System (GIS) database is a powerful tool that helps scientists and attorneys conceptualize complex environmental conditions and manage large amounts of information. A GIS database can assimilate information from multiple sources and time periods about a location or series of locations into a database structure. These types of databases also possess a spatial component that is integral to the collection, visualization, organization, analysis, and reporting of data from multiple sources (see graphic). In environmental litigation, a well constructed GIS database can help document and visualize a site's history, organize and manage information in a centralized location, and help evaluate potential contaminant exposures.

> ***GIS databases are only as good as the information on which they are based…***

Many environmental sites have complex industrial histories that can date back to the 1800s. A GIS database can be used to track the "evolution" of any location of interest in environmental litigation. Often a site and the surrounding area will undergo significant changes in land use or land cover over time. Understanding the time series variations in locations, elevations, dimensions, and characteristics of pertinent features (such as parcel boundaries, building footprints, utility lines, water bodies, or sampling locations) is critical in developing an accurate picture of how a particular area has evolved. Another common use for a GIS database is as a data repository interface. Information such as aerial or satellite images, bathymetry, historical maps, or plot and parcel plans can be cataloged and aligned within a GIS database to help evaluate the locations of historical buildings or structures, conceptualize the preferential pathways for surface and subsurface transport of chemicals, and visualize and interpolate the concentrations of chemicals in soil, air, groundwater, or surface water.

Developing and managing a centralized database can allow multiple litigation team members to access, review, and track documents and data from a consolidated, quality controlled source. This can help litigation team members or experts perform their work in parallel with one another, reduce time spent iteratively harmonizing results and conclusions, and effectively reduce costs. Creating one repository for information can help minimize duplication of documents, allow for better version tracking, and streamline the process of data sharing and document production. In addition, any available information can be linked to a location, viewed in a web-based map, or displayed as a trial demonstrative or a figure in an expert report.

Typical GIS Database Development Workflow



A GIS database can also be used to evaluate alleged plaintiff exposures in toxic tort cases by showing how a plaintiff's residences and places of employment changed over time, and how those addresses relate to an alleged pollutant source. Photos, chemical testing values or graphs, and other information can be linked to an address or location and visualized in many different ways.

There are, of course, limitations. GIS databases are only as good as the information on which they are based, and building an accurate GIS database from numerous and often disparate historical or contemporary documents can be a time-consuming process. With the massive amount of information usually available in environmental litigation, a GIS professional must assess the quality of the information available to estimate uncertainty in spatial and historical accuracy. In addition, while a GIS database possesses tools that can help to accurately interpolate between known data sets and locations, it is not typically used for predictive modeling or forecasting.

When constructed and implemented properly, a GIS database is an extremely effective tool that can be leveraged by multiple parties to collect, organize, store, query, and visualize key information in environmental litigation cases.

The authors can be reached at mmayo@gradientcorp.com and nslagowski@gradientcorp.com.

Location, Location…Association? Analyzing Spatial Relations in Environmental Epidemiology

By Christine T. Loftus, Ph.D., M.S., M.P.H.

Spatial epidemiology studies can prevent the need to gather original data, but they are not a panacea and must be used wisely.

The application of geospatial methods in environmental epidemiology dates back to Dr. John Snow, an English physician who tracked the spread of the 1854 cholera epidemic in London using simple street maps marked to indicate homes of cholera cases. By inspecting potential "clusters" of the disease across the region, Dr. Snow pinpointed a single water pump he hypothesized to be the source of contaminated water, and according to some historians, removal of the handle from this single pump dramatically diminished cholera deaths and ended the epidemic.

…many researchers urge caution in interpreting the results of spatial analyses…

More recently, the widespread availability of GIS tools and improved computing power for managing large geodatabases have fueled a resurgence in geospatial analyses in health research. For example, the annual number of articles indexed in PubMed with the keywords "spatial epidemiology" has increased dramatically from 43 to 934 between 1990 and 2014.

Strictly speaking, geospatial epidemiology encompasses any investigation of disease that accounts for subject locations – residential, occupational, or otherwise – in analysis. Three main types of geospatial analyses are commonly encountered in environmental epidemiology: disease mapping, cluster analyses, and geographic correlation studies. Both disease mapping and cluster analyses are aimed at describing spatial patterns in disease rates, and are valuable for disease surveillance as well as for generation of hypotheses about links between disease and environmental exposures. In geographic correlation studies, disease outcomes are quantitatively analyzed in associations with spatial relations, usually by means of regression. For example, several studies have evaluated whether the risk of pediatric asthma is associated with residential proximity to high-traffic roads (*i.e.*, Boehmer *et al.*, 2013).

It is critical to understand the advantages and limitations common in spatial epidemiology studies as this field of research grows rapidly in popularity. Spatial approaches to environmental epidemiology are attractive because analyses often can be conducted using existing data sets, circumventing the need for original data collection. For example, using state-maintained hospital discharge records alone, a researcher can map disease rates across a region and assess whether higher prevalence areas correspond to suspected sources of pollution. Another advantage of geospatial epidemiology is that maps of disease and potential exposure sources are visually appealing and easily capture the attention of technical and non-technical audiences alike.

On the other hand, important limitations arise from the simple assumption that location is a proxy for exposure to environmental pollutants. For example, McKenzie *et al.* (2014) investigated adverse birth outcomes in association with distance of maternal residence to hydraulic fracking wells. They explored the hypothesis that women residing closer to wells during pregnancy were more likely exposed to volatile organic compounds (VOCs) and other air pollutants emanating from well sites and subsequently at higher risk for poor birth outcomes. The use of residential distance as a surrogate of air pollutant exposure, however, disregards numerous other influences on exposure, such as wind speed, wind direction, and other meteorological factors; indoor-outdoor differences in pollutant concentrations; daily travel away from home; and local topography. Any difference between true exposure and estimated exposure (*i.e.*, exposure measurement error) can bias epidemiologic associations in a negative or positive direction.

On account of these strengths and limitations, many researchers urge caution in interpreting the results of spatial analyses, while recommending them to be used as tools to explore associations and generate hypotheses. As new results from spatial epidemiology emerge in the scientific literature daily, a critical assessment of the conclusions, uncertainties, and potential biases is warranted.

The author can be reached at cloftus@gradientcorp.com.

References:

Boehmer, T.K., S.L. Foster, J.R. Henry, E.L. Woghiren-Akinnifesi, and F.Y. Yip. 2013. Residential Proximity to Major Highways – United States, 2010. *Morbidity and Mortality Weekly Report* (MMWR) 62(3 Suppl.):46-50.

McKenzie, L.M., R. Guo, R.Z. Witter, D.A. Savitz, L.S. Newman, and J.L. Adgate. 2014. Birth outcomes and maternal residential proximity to natural gas development in rural Colorado. *Environ. Health Perspect.* 122 (4):412-417.

What's New at Gradient

Awards and Announcements

Barbara D. Beck has been recertified as an EU-Registered Toxicologist through the UK Register of Toxicologists.

Julie Goodman received the Distinguished Alumna Award from the Johns Hopkins University Alumni Association, and was an invited observer at the International Agency for Research on Cancer meeting in Lyon, France, June 2-9.

Publications

Goodman, J.E., C.T. Loftus, and **K. Zu**. 2015. 2,4-Dichlorophenoxyacetic Acid and Non-Hodgkin's Lymphoma, Gastric Cancer, and Prostate Cancer: Meta-analyses of the Published Literature. *Ann. Epidemiol.* 25(8):626-636.

Goodman, J.E., R.L. Prueitt, S.N. Sax, D.M. Pizzurro, H.N. Lynch, K. Zu, and F.J. Venditti. 2015. Ozone Exposure and Systemic Biomarkers: Evaluation of Evidence for Adverse Cardiovascular Health Impacts. *Crit. Rev. Toxicol.* 45(5):412-452.

Loftus, C.T., M. Yost, P. Sampson, G. Arias, E. Torres, V.B. Vasquez, P. Bhatti, and C. Karr. 2015. Regional PM2.5 and asthma morbidity in an agricultural community: a panel study. *Environ. Res*. 136:505-512.

Petito Boyce, C., J.E. Goodman, S.N. Sax, and **C.T. Loftus.** 2015. Providing perspective for interpreting cardiovascular mortality risks associated with ozone exposures. *Regul. Toxicol. Pharmacol.* 72(1):107-116.

Rohr, A.C., S.L. Campleman, **C.M. Long, M.K. Peterson**, S. Weatherstone, W. Quick, and **A.S. Lewis**. 2015. Potential occupational exposures and health risks associated with biomass-based power generation. *IJERPH.* 12(7):8542-8605.

Rominger, J.T. and H.M. Nepf. 2014. Effects of blade flexural rigidity on drag force and mass transfer rates in model blades. *Limnol. Oceanogr*. 59(6):2028-2041.

Sidhu, M.S., K.P. Desai, **H.N. Lynch, L.R. Rhomberg, B.D. Beck**, and F.J. Venditti. 2015. Mechanisms of action for arsenic in cardiovascular toxicity and implications for risk assessment. *Toxicology*. 331:78-99.

Young, D., **J.W. Rice**, R. Martin, E. Lindquist, A. Lipzen, I. Grigoriev, and D. Hibbett. 2015. Degradation of Bunker C Fuel Oil by White-Rot Fungi in Sawdust Cultures Suggests Potential Applications in Bioremediation. *PLoS One*. 10(6):e0130381.

Upcoming Presentations

Henderson, NV. Oct. 18-22, 2015. Annual Meeting of the International Society of Exposure Science.

- **"Assessment of Particulate Matter Air Quality Impacts and Potential Health Risks Posed by an Urban Building Demolition Project."** C.M. Long.

Amherst, MA. Oct. 19-22, 2015. Annual International Conference on Soils, Sediments, Water, and Energy.

- **"Vapor Intrusion Pathway: An Update to State Guidance Documents."** L.C. Levy.

Salt Lake City, UT. Nov. 1-5, 2015. Society of Environmental Toxicology and Chemistry NA Annual Meeting.

- **"A Comparison of Environmental Assessment Requirements of New Human Drugs in the U.S. and the EU."** T.D. Lunsman, T. Verslycke.
- **"The Crystalline Silica Conundrum in the Classification of Products Using the Globally Harmonized System for Classification and Labeling of Chemicals."** A.S. Lewis, C.B. Tuit.
- **"Degradation Products as Surrogates for Hazard Assessment of Readily Degradable Substances."** T.D. Lunsman, H.C. Ritter, D.M. Pizzurro.
- **"Ecotoxicity Monitoring of Plant-Based Biodiesel Contaminated Soil Using the Earthworm (Eisenia fetida)."** I.A. Bamgbose, T. Anderson.
- **"The Globally Harmonized System of Classification & Labeling of Chemicals (GHS): Challenges and Opportunities."** A.S. Lewis, T.A. Verslycke, J. Zhang, K.R. Reid, J.M. Kneeland.
- **"GreenScreen: Looking Beyond the Benchmark."** T.A. Lewandowski, A.S. Lewis, K.R. Reid, D.G. Skall, K.S. Sullivan, T. Verslycke, J. Zhang.
- **"An Iterative and Multidisciplinary Framework for Determining Read-Across Chemical Surrogates."** J.W. Rice, H.C. Ritter, J.M. Kneeland, J. Zhang, C. Butler, A.E. Noble.
- **"Mercury Discharge into the Ohio River in an Evolving Regulatory Environment."** C.B. Tuit, D.E. Langseth.
- **"Organic Arsenic Trends Naturally Found in Seafood."** G.I. Greenberg, I.A. Bamgbose, D.B. Mayfield.
- **"Refining Ecological Soil Screening Levels Using Benchmark Dose Analysis."** D.G. Skall, D.B. Mayfield, M. Johnson.
- **"The Regulatory Framework for Read-Across and Surrogate Use for Hazard Assessment."** J.M. Kneeland, J.M. Cohen.

Guest Editorial: Litigators and Experts: Ignore GIS Technology at Your Peril

By Christopher N. Thatch, Esq.

GIS database technology played a pivotal role in the winning defense of a company facing potential multi-billion dollar liability in Hurricane Katrina litigation.

In 2012, our client Washington Group International (WGI) faced billions of dollars in claims for allegedly causing the floodwall failures at the east bank of the Inner Harbor Navigation Canal that decimated the Lower Ninth Ward during Hurricane Katrina. The trial was long and complicated, with testimony from 14 expert witnesses, several hundred exhibits, and more details about dirt than some engineers learn in a lifetime. Ultimately, the court sided with our client and concluded that WGI's environmental and site clearing project at the east bank in no way contributed to the floodwall failures or the resulting damage to the Lower Nine. Using Geographic Information System (GIS) technology to develop a proper understanding of the geography of New Orleans was critical to WGI's victory.

The value of our GIS-driven analysis cannot be overstated.

GIS databases are valuable tools for processing and organizing large quantities of spatial or geographic data. Once placed into a GIS database, a wide variety of data types from numerous sources and time periods can be aligned, cataloged, and analyzed together to facilitate a more complete understanding of the information. GIS databases also help minimize errors in an expert's analysis and equip attorneys and experts to explain complicated facts to a jury or judge.

This was particularly true in our case. Confronted with decades of information about the geological conditions and flood protection structures in New Orleans, our experts relied on several sources to better understand the pre-storm conditions at the east bank of the Canal, including historical maps from the U.S. Army Corps of Engineers and local levee boards, geological reports from the U.S. Geological Survey, and aerial images obtained from years of survey work in the region. Using GIS technology, data from these sources was extracted, digitized, and overlaid with the results of a contemporaneous on-site soil investigation to paint a complete picture of the geography at the east bank.

The value of our GIS-driven analysis cannot be overstated. Not only did GIS technology make it possible for us to organize and examine thousands upon thousands of electronic files, images, and other materials, but it also allowed us to describe with precision how Hurricane Katrina impacted the subsurface conditions at the east bank – an exercise that even the Army Corps' foremost geotechnical engineers found challenging in the wake of the storm. Armed with this information, we were able to accurately portray the relevant facts, discredit the opposing experts' science, and present an engaging and persuasive narrative that, in the end, the court agreed with.

In this ever changing and increasingly complex world, GIS technology will continue to be a necessary tool for collecting and examining data in litigation.

Christopher N. Thatch is a trial attorney at the law firm Jones Day in Washington, D.C., and a member of the firm's Business & Tort Litigation Practice. The case that is the subject of this editorial is In re Katrina Canal Breaches Consolidated Litigation, No. 05-4182 (E.D. La.). The views expressed in this editorial are the personal views of the author and do not necessarily reflect those of Jones Day. The author can be reached at cthatch@jonesday.com.

By The Way...

Gradient is now one of four companies globally that are certified GreenScreen® Profilers.

Gradient
20 University Road
Cambridge, Massachusetts 02138
(617) 395-5000
trends@gradientcorp.com
www.gradientcorp.com

The next issue will focus on:
Agricultural Chemicals

Trends is a free publication of Gradient. If you would like to be added to the distribution list, email trends@gradientcorp.com.

Look for *Trends* on the Web at: www.gradientcorp.com