

Scientists and Librarians Create an Environmental Toxicology Data Repository

Deborah Keil^I and Kenning Arlitsch^{II}

^I Department of Pathology, University of Utah, ^{II} University of Utah J. Willard Marriott Library

This paper proposes the development of an environmental toxicology data repository for the Wasatch Front, the major metropolitan area in the state of Utah. This repository, developed by scientists and librarians, will address gaps in national-level repositories that cannot take into account unique local issues caused by geography, industry, military installations, and population. The repository will enable local clinicians and academic researchers to collect data, describe it using a standard metadata schema, preserve it for the future, and make it accessible through common search engines.

Introduction

The Wasatch Front in Utah encompasses several cities in north central Utah and is home to most of the state's population. The area is characterized by its rugged beauty and topographical contrasts, with high, snow-capped mountains to the east transitioning to the Great Basin Desert in the west. The Wasatch Mountain range rises more than 1,500 meters from the valley floor in Salt Lake City to heights of 3,600 meters, making it one of the most precipitous elevation gains in proximity to a major urban area. The Great Salt Lake is another dominant feature of the landscape, providing a migratory stopover for birds as well as supporting a multi-million dollar brine shrimp industry, the only life the lake supports. The Great Basin is a vast area of alternating

Tokar, A., Beurskens, M., Keuneke, S., Mahrt, M., Peters, I., Puschmann, C., van Treeck, T., & Weller, K. (Eds.). (2012). *Science and the Internet* (pp. 145-155). Düsseldorf: Düsseldorf University Press

desert basins and high mountain ranges that stretches across Nevada to the Sierra Mountains in California. One of its defining characteristics is that there are no significant rivers and none of its water escapes to either ocean. Eighty percent of Utah's human population—just over two million people—lives along the Wasatch Front, and this population depends on the 500 inches of annual average snowfall on the mountains that produces life-giving water. (“Wasatch Front,” 2012).

The writer Wallace Stegner (1960) referred to the American West as the “geography of hope” but also warned of its fragility under man's heavy hand. The geography of the Wasatch Front, which is so stunning, also creates environmental conditions that exacerbate the effects of man-made pollution and, consequently, affect human health. The topography creates a natural phenomenon known as temperature inversions, where warm air at the elevation of the mountains presses cold air to the valley floors during winter high-pressure systems, trapping the emissions of industry, automobiles, and homes and sometimes creating some of the worst air quality in the nation (Garber, 2011). The presence of major extractive industries, military installations, and a rapidly growing population all contribute to environmental pollutants. Air, water, and soil are compromised, and the long-term health effects of these conditions remain unclear. Some health conditions may be correlated: The region has been identified as exceeding national population averages in the incidence of asthma (May, 2011; KSL, 2007) and autism (Baio, 2012), while statistically significant increases in lung and bronchial cancer were identified near Hill Air Force Base (Ball et al., 2008).

The Wasatch Front is home to nationally renowned medical institutions, the Utah State Health Department and world-class universities with researchers who investigate environmental pollutants and related health issues. However, there is no local infrastructure to facilitate the exchange of information among these regional experts and institutions. The identification of potential cause-and-effect relationships of illnesses due to environmental chemical contaminants depends on the availability of reliable human health and exposure data. Developing a cooperative infrastructure for sharing and combining data could identify local themes and prioritize action according to regional environmental health data trends.

Research Problem

Most environmental toxicology data repositories have a national scope and fail to consider the unique regional effects of local topography, industry, and population. This paper proposes an environmental toxicology data repository

focused on the Wasatch Front that could function as a model for other regionally based repositories. The goal is to merge toxicologically relevant and human health data in a regional centralized repository. Such a data repository on a flexible, interoperable, and accessible platform would provide for the development of computational risk assessment models to address regional environmental insults and improve health outcomes.

Potential Sources of Environmental Contaminants on the Wasatch Front

Extractive Industries

One of the largest open-pit copper mines in the world has been in operation in the Salt Lake valley since the early part of the twentieth century and has fuelled a significant part of the area's economy (Arrington & Hansen, 1963). Open-pit hard rock mining is environmentally destructive and produces large amounts of waste (known as tailings) because only a small portion of the total mined material contains the sought-after ore. The tailings contain heavy metals such as lead, zinc, and cadmium and can contaminate the soil, creating potential health hazards for humans and wildlife. Vegetables grown in mine tailings have been shown to accumulate heavy metals in the edible parts of the plants (Cobb et al., 2000). Additionally, metal smelters release gases such as carbon dioxide (CO₂), sulphur dioxide (SO₂), and nitrogen oxide (NO_x) (Dudka & Adriano, 1997).

On the west side of the Great Salt Lake is U.S. Magnesium Corporation, formerly known as MagCorp. The company was sued by the U.S. Department of Justice (DOJ) on behalf of the Environmental Protection Agency (EPA), and the DOJ (2001) noted that for

many years MagCorp has ranked Number One on the EPA's toxic release inventory, based on its chlorine emissions.

The federal government sued again in 2005, alleging that the plant was "the nation's worst polluter" and that it "illegally manufactured and dumped carcinogenic polychlorinated biphenyls, or PCBs" (Henetz, 2005). MagCorp filed for bankruptcy, and its subsequent restructuring as U.S. Magnesium helped it to rebuff the lawsuits. U.S. Magnesium has made substantial improvements in its emissions, though it is still considered one of the top five polluters in Utah (Fahys, 2008).

Military Installations

The Wasatch Front is home to several military installations, such as the Dugway Proving Ground and the Tooele Army Depot in the desert to the west of Salt Lake City and Hill Air Force Base to the north.

The U.S. Army manages the Dugway Proving Ground, and its mission is to test biological and chemical weapon systems. Dugway also serves as a training ground for fighter pilots from Hill Air Force Base. In 1968, six thousand sheep died in Skull Valley, thirty miles from Dugway, due to apparent exposure from an accidental release of the VX nerve agent (Boffey, 1968). More recently, scientists at Dugway have worked with anthrax spores (Matsumoto, 2003).

The Tooele Army Depot once stored approximately 42% of the United States' chemical weapons stockpile, including sarin, mustard, and VX blister and nerve agents, some of which "could persist in the environment long after an accidental release" (Carnes & Watson, 1989). While these weapons are mainly known for their acute lethality, the "possibility of long-term brain dysfunction after exposure to a nerve agent has also been raised" (Carnes & Watson, 1989). In 1991, President George H. W. Bush reversed decades-old U.S. policy by "foreswearing the use of chemical weapons for any reason," effectively eliminating the need to maintain a stockpile of these weapons (Foote, 1994). As a result of this policy change and because of the dangers of the long-term storage of these aging weapons, a plan to incinerate them was developed, and by 2012, the chemical weapons stockpile was reduced by 99%.

Hill Air Force Base is located in Ogden and is one of the largest employers in Utah. It has been listed as a Superfund site by the Environmental Protection Agency since 1987 (U.S. Environmental Protection Agency, 2012) due to soil contamination from polychlorinated biphenyls (PCBs) and a degreasing solvent known as trichloroethylene (TCE). While the PCBs have largely been removed from the soil, the TCE contamination has spread into the aquifer adjacent to the base and continues to migrate as a subterranean plume (Jackson & Dwarakanath, 1999). TCE is a known carcinogen, and the cleanup is expected to last another 65 years (Dougherty, 2010).

Home and Automobile Emissions

The population density along the Wasatch Front can be expected to contribute to particulate emissions. Home furnaces and fireplaces, as well as gasoline and diesel motor vehicles produce pollution that adds to industrial pollution and affects air quality, adding to the industrial production of particulate

matter (PM) (Utah Department of Environmental Quality: Division of Air Quality, 2012). Wood smoke from natural or human-started wildfires also add particulate pollution (Long et al., 2002), which can vary in size from fine (2.5 microns and smaller diameter) to coarse (10 microns). Fine PM can pass deep into the lungs and has been linked to cardiovascular disease and specifically to heart disease events triggered by short-term exposure along the Wasatch Front (Pope et al., 2006). Particulate air pollution increases mortality in the region (Pope et al., 1999).

Human Biomonitoring

Population-based biomonitoring programs in the United States began with the efforts of the National Human Monitoring Program (NHMP), administered by the U.S. Environmental Protection Agency, and National Health and Nutrition Examination Survey (NHANES), administered by the United States Centers for Disease Control and Prevention. In the last decade, the National Human Exposure Assessment Survey (NHEXAS), administered by the U.S. Environmental Protection Agency, and NHANES (1999-2000) have expanded the list of chemicals measured to upward of 140 (National Research Council (U.S.) Committee on Human Biomonitoring for Environmental Toxicants, 2006). This process provides valuable data about background exposure levels and trends in the population as a whole. However, it is less effective for addressing local issues, and it does not consider environmental monitoring data.

The U.S. is geographically expansive with great population diversity, and each region may be defined by unique health concerns. It would seem reasonable to implement a local biomonitoring project to identify unique local issues and to address health problems. The coordination of a local surveillance program may provide for more agility in human biomonitoring, as the approach would be scaled to greater responsiveness for local environmental concerns. Environmental issues may be uniquely assessed for the community by combining local human biomonitoring data, hospitalization data, and environmental contaminant sources. Based on current toxicological data combined with an expanding population located near a number of extractive industries, military installations, and unique geologically restrictive formations, the potential risks to human health call for a Wasatch Front regional human biomonitoring program.

Within the scope of environmental toxicology, it is increasingly important to collect exposure information and articulate this with corresponding human health. The National Research Council (NRC) of the National Academy of

Science (NAS) emphasizes the importance of collecting human exposure, population susceptibility, environmentally relevant hazard, and biomonitoring data (National Research Council Committee on Toxicity Testing and Assessment of Environmental Agents, 2007). Although the literature characterizing environmental chemicals in controlled laboratory studies has grown substantially, the primary limitations in human health risk assessment lie in linking this laboratory-based data to real-world exposures.

The Wasatch Front serves as an excellent model for such an undertaking. Health disparities include increased rates of autism, asthma, and lung and bronchial cancers, and this disease pattern differs from those identified nationally. By identifying these local trends, unique health priorities for this region can be established. Sharing data and conclusions with policy and health professionals facilitates the ability and the urgency to address local issues.

Data Management

The rapidly accumulating literature on data management points to significant challenges and opportunities for scientists and librarians working collaboratively. Health- and science-related disciplines are generating an unprecedented deluge of data, and the scientific community is struggling to harness larger data sets and increasingly complex data in a transparent and integrated fashion. Currently, some exposure assessment data that identifies sources and chemicals is housed in government organizational “siloes” such as the National Center for Environmental Assessment (NCEA), administered by the U.S. EPA, National Exposure Research Laboratory (NERL), administered by the U.S. EPA, and Regional EPAs. The challenge is to create interoperability between these repositories to more effectively prioritize environmental health threats. Achieving multi-scale integrations is necessary to advance translational science.

Recently, National Institute of Environmental Health Sciences (NIEHS), administered by the U.S. National Institutes of Health, has made available Chemical Effects in Biological Systems (CEBS), administered by the U.S. NIH, the first public repository for toxicology data. CEBS was originally developed in 2002 to house genomic data and was recently expanded to include experimental design with animal, human, and cell culture data in toxicology. Adapting or linking this model to a local repository may maximize local information for a community.

Challenges

Most scientists struggle to manage the data sets that support their research, particularly as the ability to produce data increases. Numerous barriers to effective management have been described; some are technical in nature, while others are non-technical. Technical barriers include storage, metadata, and software, while non-technical barriers can include fear of competition, lack of trust, lack of incentives, lack of control (Feijen, 2011), and concerns about data quality (Research Information Network & National Endowment for Science Technology and the Arts, 2010).

Librarians must also confront the challenges of providing high-quality data management services. A unique set of skills is required to address the finer points of repository management, discipline-specific metadata, search engine optimization, and digital preservation; many librarians are unprepared.

Opportunities

There are obvious advantages that come from managing data in a formalized structure, including minimizing loss and maximizing retrieval and reuse. Recent research also demonstrates that authors who make available the data sets that support their publications enjoy a higher rate of citations (Piwowar et al., 2007). Making data available has been characterized as necessary to advance scientific research and solve global problems by some authors (Faniel & Zimmerman, 2011), as well as promoting efficiency in research and scholarly rigor by others (Research Information Network & National Endowment for Science Technology and the Arts, 2010).

There are many advantages to combining toxicology and health data in a central or coordinated repository. The interpretation of a study may be broadened when considered in the perspective of related studies, and new relationships or queries may be generated. As the data are combined in a flexible format within an institutional repository, a future foundation for computational modeling may be established. Currently, limited computational models are available to query between data sets. As technology advances, these data are poised for utilization in future discoveries and applications of risk assessment.

Scientists' Data Management Needs

Data management needs for most scientists can be divided into two broad areas: working data and archival data. Working data are shared and used by co-researchers as they conduct their research. Access to the data is generally limited to members of the team, and data at this stage is closely guarded. The

technology needs include restricted workspaces that allow team members to share files and version control features. It may also include the need for electronic laboratory notebook software.

Data sets become archival after researchers have published their findings and agreed to release all or a subset of their data. Many universities have created institutional repositories to permanently archive the scholarly publications of their researchers, and institutional repositories may be appropriate for data sets as well.

Wasatch Front Repository Proposal

The environmental toxicology repository proposal for the Wasatch Front consists of two parts: 1) a repository run by the University of Utah built on an existing infrastructure, and 2) inclusion of that repository (as well as others) in the Mountain West Digital Library (MWDL).

The University of Utah manages an institutional repository known as USpace (<http://uspace.utah.edu>), whose mission is to capture the intellectual output of the University of Utah faculty, staff, and students. USpace consists of several collections, such as electronic theses and dissertations, scholarly papers, and university administrative records. Adding another collection for data sets is feasible.

Other researchers, particularly those unaffiliated with the University of Utah may not wish to store their data in a university-owned repository. The MWDL was founded in 2002 under the auspices of the Utah Academic Library Consortium and is a distributed digital library portal that harvests only metadata (not objects) via Open Archives Initiative—Protocol for Metadata Harvesting (OAI-PMH) from approximately eighteen remote hosting sites, mostly academic libraries. Those sites, in turn, support the digital collections of dozens of partners—many of which are not libraries—that lack their own infrastructure. Search results in the MWDL portal link users to the hosting site where the particular object or collection exists; this practice allows owning institutions and researchers to maintain the control and identity of their collections. The MWDL has a formalized partnership agreement and a metadata application profile.

Scientists wishing to make available their toxicology data and publications related to Wasatch Front environmental conditions could have the option of submitting to USpace or to another repository that is OAI-PMH-compliant and adheres to the MWDL Dublin Core Application Profile (Walters et al., 2011). Metadata describing the data sets would be harvested into the aggregated MWDL index hosted at the University of Utah, but the data sets and publications themselves would remain with the hosting site.

Conclusion

The geography of the Wasatch Front compounds the effects on human health of population growth, as well as industrial and military environmental pollution. National environmental toxicology data repositories do not consider the unique local characteristics that appear to have a detrimental effect on air and soil quality and may correlate to higher incidences of asthma, autism, and other environmental health-related problems.

The creation of a local environmental toxicology data repository may help to address some of the serious health issues and may serve as a model for other regionally focused data repositories. The established political and technical infrastructures of the USpace institutional repository at the University of Utah and the regional MWDL can be leveraged to quickly establish this repository. The approach is intended to be economical and adaptable to other regional areas that are supported by an academic library. Librarians themselves may recognize this proposal as an opportunity for greater involvement in data management and to work more closely with scientists and other researchers.

References

- ARRINGTON, L. J., & HANSEN, G. B. (1963). *The richest hole on earth: A history of the Bingham Copper Mine*. Logan, UT: Utah State University Press.
- BAIO, J. (2012). Prevalence of autism spectrum disorders—Autism and developmental disabilities monitoring network. *Morbidity and Mortality Weekly Report*, 61(3), 1-19.
- BALL, W., LEFEVRE, S., JARUP, L., & BEALE, L. (2008). Comparison of different methods for spatial analysis of cancer data in Utah. *Environmental Health Perspectives*, 116(8), 1120-1124.
- BOFFEY, P. M. (1968). Nerve gas: Dugway accident linked to Utah sheep kill. *Science*, 162(3861), 1460-1464.
- CARNES, S. A., & WATSON, A. P. (1989). Disposing of the US chemical weapons stockpile: An approaching reality. *The Journal of the American Medical Association*, 262(5), 653-659.
- COBB, G. P., SANDS, K., WATERS, M., WIXSON, B. G., & DORWARD-KING, E. (2000). Accumulation of heavy metals by vegetables grown in mine wastes. *Environmental Toxicology and Chemistry*, 19(3), 600-607.
- DOUGHERTY, J. M. (2010, March 28). No further action needed in Hill Air Force Base PCB cleanup. *Deseret News*. Retrieved September 29, 2012 from

- <http://www.deseretnews.com/article/700018954/No-further-action-needed-in-Hill-Air-Force-Base-PCB-cleanup.html?pg=all>
- DUDKA, S., & ADRIANO, D. C. (1997). Environmental impacts of metal ore mining and processing: A review. *Journal of Environment Quality*, 26(3), 590.
- Experts puzzled over some high numbers in Utah asthma report. (2007, June 11). *The Salt Lake Tribune*. Retrieved November 10, 2012 from <http://bit.ly/RP8QvN>
- FAHYS, J. (2008). EPA: U.S. Magnesium wastes endangers workers, families, birds. *Salt Lake Tribune*. Retrieved November 10, 2012 from <http://tinyurl.com/bearqwl>
- FANIEL, I. M., & ZIMMERMAN, A. (2011). Beyond the data deluge: A research agenda for large-scale data sharing and reuse. *International Journal of Digital Curation*, 6(1), 58-69.
- FEIJEN, M. (2011). What researchers want. *SURF*, 57(2), 1-32.
- FOOTE, W. (1994). The chemical demilitarization program—Will it destroy the nation's stockpile of chemical weapons by December 31, 2004? *Military Law Review*, 146, 1-93.
- GARBER, H. (2011). Air quality on the Wasatch Front is a public health crisis. *Utah Physicians for Healthy Environment*. Retrieved November 10, 2012 from <http://tinyurl.com/b5cs885>
- HENETZ, P. (2005, May 14). Feds sue plant for alleged PCB spill. *Salt Lake Tribune*. Retrieved September 29, 2012 from http://www.sltrib.com/ci_2734847
- JACKSON, R. E., & DWARAKANATH, V. (1999). Chlorinated decreasing solvents: Physical-chemical properties affecting aquifer contamination and remediation. *Ground Water Monitoring & Remediation*, 19(4), 102-110.
- LONG, R. W., SMITH, R., SMITH, S., EATOUGH, N. L., MANGELSON, N. F., EATOUGH, D. J., POPE, C. A., & WILSON, W. E. (2002). Sources of fine particulate material along the Wasatch Front. *Energy & Fuels*, 16(2), 282-293.
- MATSUMOTO, G. (2003). Anthrax powder: State of the art? *Science*, 302(5650), 1492-1497.
- MAY, H. (2011, March 11). Asthma: Why do some Utah kids struggle harder to breathe? *Salt Lake Tribune*. Retrieved September 30, 2012 from <http://www.sltrib.com/sltrib/news/51289555-78/healthyforwhom-asthma-utah-percent.html.csp>
- NATIONAL RESEARCH COUNCIL (U.S.) COMMITTEE ON HUMAN BIOMONITORING FOR ENVIRONMENTAL TOXICANTS (2006). *Human biomonitoring for environmental chemicals*. Washington, D.C.: National Academies Press.
- NATIONAL RESEARCH COUNCIL COMMITTEE ON TOXICITY TESTING AND ASSESSMENT OF ENVIRONMENTAL AGENTS (2007). *Toxicity testing in the 21st century: A vision and a strategy*. Washington, D.C.: National Academies Press.
- PIWOWAR, H. A., DAY, R. S., & FRIDSMA, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS one*, 2(3), e308.

- POPE, C. A., HILL, R. W., & VILLEGAS, G. M. (1999). Particulate air pollution and daily mortality on Utah's Wasatch Front. *Environmental Health Perspectives*, 107(7), 567-573.
- POPE, C. A., MUHLESTEIN, J. B., MAY, H. T., RENLUND, D. G., ANDERSON, J. L., & HORNE, B. D. (2006). Ischemic heart disease events triggered by short-term exposure to fine particulate air pollution. *Circulation*, 114(23), 2443-2448.
- RESEARCH INFORMATION NETWORK & NATIONAL ENDOWMENT FOR SCIENCE TECHNOLOGY AND THE ARTS (2010). *Open to all? Case studies of openness in research*. Retrieved September 29, 2012 from <http://apo.org.au/research/open-all-case-studies-openness-research>
- STEGNER, W. (1960). Wilderness letter. Retrieved September 29, 2012 from <http://www.colorado.edu/AmStudies/lewis/west/wilderletter.pdf>
- U.S. DEPARTMENT OF JUSTICE (2001). U.S. files suit against Magnesium Corporation of America for toxic pollution near Great Salt Lake. Retrieved September 29, 2012 from <http://www.justice.gov/opa/pr/2001/January/021enrd.htm>
- U.S. ENVIRONMENTAL PROTECTION AGENCY (2012). Superfund program—Hill Air Force Base. Retrieved November 10, 2012 from <http://tinyurl.com/avk7tvd>
- UTAH DEPARTMENT OF ENVIRONMENTAL QUALITY: DIVISION OF AIR QUALITY (2012). Air pollutants: Particulate matter. Retrieved September 29, 2012 from http://www.airquality.utah.gov/Public-Interest/about_pollutants/about_pm.htm
- WALTERS, C., WILLEY, K., KELLER, T. W., LAMPERT, C., MATTHEWS, R., MCINTYRE, S., MOWER, A., RIDING, E., ROBERTSON, J., SOUTHWICK, S., STRACK, G., & WOOLCOTT, L. (2011). *Mountain West Digital Library Dublin core application profile*. Retrieved November 10, 2012 from <http://tinyurl.com/a488nyn>
- Wasatch Front. (2012, September 25). In *Wikipedia*. Retrieved September 29, 2012 from http://en.wikipedia.org/wiki/Wasatch_Front