



HHS Public Access

Author manuscript

Front Sustain Food Syst. Author manuscript; available in PMC 2022 June 04.

Published in final edited form as:

Front Sustain Food Syst. 2021 June ; 5: . doi:10.3389/fsufs.2021.620470.

Alleviating Environmental Health Disparities Through Community Science and Data Integration

Mónica D. Ramírez-Andreotta^{1,2,*}, Ramona Walls³, Ken Youens-Clark⁴, Kai Blumberg⁴, Katherine E. Isaacs⁵, Dorsey Kaufmann¹, Raina M. Maier¹

¹ Department of Environmental Science, University of Arizona, Tucson, AZ, United States

² Mel and Enid Zuckerman College of Public Health's Division of Community, Environment and Policy, University of Arizona, Tucson, AZ, United States

³ BIO5 Institute, University of Arizona, Tucson, AZ, United States

⁴ Department of Biosystems Engineering, University of Arizona, Tucson, AZ, United States

⁵ Department of Computer Science, University of Arizona, Tucson, AZ, United States

Abstract

Environmental contamination is a fundamental determinant of health and well-being, and when the environment is compromised, vulnerabilities are generated. The complex challenges associated with environmental health and food security are influenced by current and emerging political, social, economic, and environmental contexts. To solve these “wicked” dilemmas, disparate public health surveillance efforts are conducted by local, state, and federal agencies. More recently, citizen/community science (CS) monitoring efforts are providing site-specific data. One of the biggest challenges in using these government datasets, let alone incorporating CS data, for a holistic assessment of environmental exposure is data management and interoperability. To facilitate a more holistic perspective and approach to solution generation, we have developed a method to provide a common data model that will allow environmental health researchers working at different scales and research domains to exchange data and ask new questions.

We anticipate that this method will help to address environmental health disparities, which are unjust and avoidable, while ensuring CS datasets are ethically integrated to achieve environmental justice. Specifically, we used a transdisciplinary research framework to develop a methodology to integrate CS data with existing governmental environmental monitoring and social attribute data (vulnerability and resilience variables) that span across 10 different federal and state agencies. A

This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

*Correspondence: Mónica D. Ramírez-Andreotta mdramire@arizona.edu.

AUTHOR CONTRIBUTIONS

MR-A: conceptualized the study and designed it with RW and writing-original draft preparation. MR-A and RW: investigation, supervision, and project administration. RW, KY-C, KB, MR-A, and KI: methodology. RW, KY-C, and KB: data curation. DK: visualization. RM, RW, MR-A, KB, and KI: writing-review and editing. RM (administrator), MR-A, and RW: funding acquisition. All authors contributed to the article and approved the submitted version.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

key challenge in integrating such different datasets is the lack of widely adopted ontologies for vulnerability and resiliency factors. In addition to following the best practice of submitting new term requests to existing ontologies to fill gaps, we have also created an application ontology, the Superfund Research Project Data Interface Ontology (SRPDIO).

Keywords

citizen science; community science; interoperability; FAIR principles; environmental health; community resiliency

INTRODUCTION

Research Context

Pollution is now the leading global cause of premature death and disease (Landrigan et al., 2018). This crisis is currently being addressed through environmental monitoring and public health surveillance efforts that are conducted by local, state, and federal agencies. Most states have environmental quality and health departments, which have the major responsibility for environmental protection and the health and safety of the population. The U.S. federal government has a number of overarching environmental and public health agencies, including the Centers for Disease Control and Prevention, the National Institutes of Health (NIH), the Food and Drug Administration, the Environmental Protection Agency (USEPA), Geological Survey (USGS), and Department of Agriculture (USDA). In addition, two other entities can play important roles in environmental quality and public health; non-governmental organizations (NGOs) and universities, both public and private, that receive extramural funding to conduct environmental quality and public health research.

One of the challenges with these efforts is that the datasets generated by each group are independent and siloed from one another, leading to a lack of standardization, interoperability, application of FAIR (Findable, Accessible, Interoperable, and Reusable) principles of data management, and stewardship (Wilkinson et al., 2016). A second challenge is that community members are rarely involved in environmental monitoring projects. Professionally paid researchers are missing key opportunities to partner with vulnerable communities, collect high resolution data, and incorporate potential exposure routes that may otherwise be overlooked (e.g., Garcia et al., 2013; Ramírez-Andreotta et al., 2013a,b; Ramírez-Andreotta et al., 2014; Dhillon, 2017; Manjón et al., 2020).

Public Participation in Scientific Research (PPSR) efforts such as citizen and community science programs (referred to as CS hereafter) can be used to address the latter challenge. PPSR is broadly defined as partnerships between scientists and non-scientists in which authentic data are collected, shared, and analyzed (e.g., Shirk et al., 2012). Such efforts have dramatically increased in the past few years (Pocock et al., 2017), and it is anticipated that this approach will permanently change the face of how scientific data are collected and who collects it. Incorporation of CS into research efforts has exciting potential due to the vast amount of data and observations that can be collected by the general public. What is most remarkable about this methodology, is the potential to redistribute power, democratize

science and achieve environmental justice (Ottinger, 2010; Pandya, 2012; Allen, 2018). CS efforts are increasingly being directed toward environmental monitoring and will be key and necessary to fully understanding the environmental determinants of chronic disease. Such monitoring information will provide the scientific basis for future prevention of environmental exposures and motivating action (Morello-Frosch et al., 2009).

The critical obstacle to using CS data in assessment of environmental exposure is data management and interoperability. As laid out in the 2016 report, “Stakeholder Analysis: International Citizen Science Stakeholder Analysis on Data Interoperability” there is empirical evidence for the importance of data standards in CS, most noteworthy is that some authorities may not use CS data because of “uncertainty about data quality assurance and quality control measures, and a lack of data standardization practices” (Gobel et al., 2016). Yet studies have confirmed that CS models can provide accurate and reliable data (e.g., Haklay, 2010; Gollan et al., 2012; Nagy et al., 2012; Tregidgo et al., 2013; Hecker et al., 2018). In order to move these data beyond disciplinary and stakeholder boundaries, data management and quality assurance is required (Haklay, 2017; Hecker et al., 2018), along with internal support and tools to effectively address the problems identified by CS. For example, the USEPA is supporting CS projects and has generated quality assurance guidance documents that include templates and handbooks to inform community members and other federal and state agencies (USEPA, 2020).

The scarcity of FAIR data (Wilkinson et al., 2016) in CS is not only unfortunate for the progression of science, but unethical. Thousands of people are contributing/participating to CS programs and co-generating datasets; dedicating their time and resources hoping that their efforts will create change and positive social-ecological outcomes (e.g., Shirk et al., 2012; Ramirez-Andreotta et al., 2015). The lack of data standardization and application of FAIR principles in government-, NGO-, university-, and CS-based efforts slows down the ability and efficiency to address environmental health disparities. Further, most public health environmental health monitoring efforts use an epidemiological approach, but it is known that epidemiology alone cannot adequately detect the effects of toxic exposures on human health (Brown, 1992; Brown and Mikkelsen, 1997). Specifically, a fundamental and critical challenge that exists in environmental justice communities is the need to account for interrelated effects of culturally-diverse and economically-disadvantaged groups with toxic exposures. Another challenge is accounting for community resiliency: the sustained ability of a community to withstand and recover from adversity (Plough et al., 2013). Community resiliency comprises the enduring capacity of geographically, politically, or affinity-bound communities to define and account for their vulnerabilities and develop capabilities to prevent, withstand, or mitigate (Abramson et al., 2015).

We hypothesize that FAIR principles can be applied to facilitate the seamless integration of CS/government/NGO/university datasets and allow the inclusion of both community vulnerabilities and resiliencies in the environmental health assessment process. This approach will allow incorporation of all viable data as well as scaling of datasets, a process that can be expected to increase the efficiency and impact of public health intervention efforts.

To address this hypothesis, we have developed a methodology to make environmental health CS data FAIR. This methodology: (1) integrates CS environmental monitoring data with other data sets to enhance discoverability and reuse of data for research translation and (2) enables better hypothesis generation. An anticipated result of this integration effort is that it will help determine if and how community-level resiliencies may combat environmental health vulnerabilities. In this methodology, we use ontologies to combine a CS dataset with existing governmental environmental monitoring and community resiliency data. An ontology is a formal specification of the concepts in a domain and the relationships among them (Gruber, 1993). The use of ontologies is a key component of FAIR data, because ontologies can transform free-text descriptions into structured, standardized machine-readable data, improving findability, interoperability, and reusability.

The CS dataset used in this research is Gardenroots, a co-created CS program. Gardenroots sees gardens as hubs for environmental health research and literacy with the goals of: engaging community members in the environmental monitoring and exposure science process; evaluating environmental quality (water, soil, and homegrown vegetables) and potential exposure routes; and designing personalized and community-based data sharing experiences to support environmental action and decision-making (Ramírez-Andreotta et al., 2013a,b; Ramírez-Andreotta et al., 2015; Sandhaus et al., 2019). CS programs such as Gardenroots demonstrate how community-engaged environmental monitoring efforts have informed local food gardening practices. By working together to determine soil quality and contaminant concentrations, Gardenroots helps sustain community and home gardens efforts while reducing chemical exposures. This is critical because community and home gardening efforts help address social and economic constraints on health by increasing access to wholesome foods, improving community building efforts, enhancing emotional well-being, creating green space, and reducing the cost of food (Ness and Powles, 1997; Armstrong, 2000; Teig et al., 2009; Ramírez-Andreotta et al., 2019). Gardenroots builds on individual- and community-level resiliencies and combats environmental health vulnerabilities, helping to ensure pollution does not interfere with local gardening efforts. However, Gardenroots is site-specific. We use this dataset to demonstrate the possibility of integration of these data with other state and federal datasets related to soil quality, food production, health, etc. This integration not only increases the spatial resolution and understanding of pollution, but also has the potential to increase environmental health decision-making capacity.

MATERIALS AND EQUIPMENT

About Gardenroots

Gardenroots was established in 2010 in collaboration with a rural community neighboring a USEPA National Priorities List site under the Superfund program slated for cleanup due to uncontrolled hazardous waste (Ramírez-Andreotta et al., 2013a,b). The site comprises a large mine tailings pile that has been barren and subject to wind and water erosion since the 1960s as well as a closed smelter facility site. Both the mine tailings and the smelter site are contaminated with high levels of arsenic, lead, and zinc (USEPA, 2013). More recently, based on the results of a community needs assessment and ongoing community engagement in Arizona, Gardenroots was continued in summer 2015 and 2019 to help address additional

community concerns regarding their soil, water, and/or plant quality. Since inception, Gardenroots has been implemented in nine communities nationwide and in AZ alone, more than 120 participants have been trained. Each Gardenroots participant completed a 2-h training on how to properly collect samples from a self-selected area. Community recruitment, trainings, and retention procedures have been previously described (Ramírez-Andreotta et al., 2015; Sandhaus et al., 2019). Typical locations included residential areas, community or school gardens, and local farms. Participants collected water, soil (yard and garden), and/or edible plant samples and submitted them to a centralized location for transport to the University of Arizona (UA). The dataset set reported here is from the following Arizona counties: Apache, Cochise, Greenlee, Pinal (Superior), and Yavapai (Dewey-Humboldt) (Figure 1). Each sample submitted and included in this dataset was analyzed for aluminum, arsenic, barium, beryllium, cadmium, chromium, copper, lead, manganese, nickel, and zinc concentrations in water (micrograms per liter, μg^{-1}), soil (milligrams per kilogram, mg kg^{-1}), and/or plant samples (mg kg^{-1}). Field and laboratory methodologies have been previously described (Ramírez-Andreotta et al., 2013a,b; Manjón et al., 2020). All Gardenroots participants received their data (individual and aggregated) via visually-rich results booklets distributed at data sharing and community gathering events or by mail (Ramírez-Andreotta et al., 2015; Sandhaus et al., 2019). In this Methods paper, the Gardenroots CS data is being used as an example to generate a methodology for others to use and allow the seamless integration of other CS collected data with existing state and federal agency datasets.

Data Management Materials and Equipment

We are using CyVerse (Merchant et al., 2016; <https://cyverse.org/>) as the primary data storage platform. CyVerse allows all project members to access and analyze data from a shared directory, thus reducing the risk of forking (having multiple, divergent copies of the same dataset). Python code for data cleanup and processing are hosted on GitHub at https://github.com/UA-SRC-data/data_loaders. The combination of shared storage and public code allows us to track exactly what processing steps were carried out on each dataset and allows others to reproduce our results. Details on the usage of these platforms is included in section Integrating CS and Federal and State Data Sources.

METHODS

Federal and State Datasets

In addition to the Gardenroots CS dataset, data were pulled from existing state and federal programs (Tables 1, 2). These datasets were selected to provide a comprehensive understanding of the possible vulnerabilities and resiliencies in Arizona rural, with special attention on medically-underserved communities that neighbor resource extraction activities. With an understanding of the possible vulnerabilities and resiliencies, efforts will be placed on gathering and juxtaposing variables to see for example, where a community has a tremendous amount of resiliency that has not been tapped for sustainability, environmental quality, and/or justice purposes or vice versa, where an area is suffering and community capacity efforts are in need.

Vulnerability Datasets

To determine vulnerability (function of the exposure and sensitivity of system, e.g., Cutter, 1996; Adger and Neil Adger, 2006; Cutter et al., 2008), datasets selected include (Table 1):

- Quality of Environment
- Quality of Health
- Social Attributes
- Access to Food.

Community Resilience Datasets

Abramson et al. (2015) proposed the Resilience Activation Framework, a conceptual model of how access to social resources promote adaptation and rapid recovery within individuals and communities. This framework rests on six described principles and assumes that access to social services can activate resilience characteristics that are inherent in both individuals and communities, and that once activated, lead to better mental and physical health and well-being (Abramson et al., 2015). Using this proposed design, community resiliencies were collected from diverse datasets and are divided here into (Table 2):

- Economic Capital
- Human Capital
- Social Capital
- Political Capital.

Integrating CS and Federal and State Data Sources

Data Processing SOP—To maximize the FAIRness of the data collected and analyzed as part of this project, we established a standard operating procedure (SOP) for all datasets that stores raw and processed data in shared folders, tracks all data processing steps, standardizes variables to existing ontologies wherever possible, and publishes standardized data to trusted repositories. Individual technologies in this SOP could be replaced with others of similar functionality. The full SOP is available at https://github.com/UA-SRC-data/data_loaders/blob/master/README.md, but in brief it describes how to:

1. Ensure a copy of the raw data is preserved and sufficiently documented: Gather raw data and store data in a shared CyVerse folder under use case name, under “raw-data.” Include a readme file to readme in each raw folder with the link to the data source and a data dictionary defining variables if needed. Also document data sources in the readme file in the appropriate directory in the data_loaders code repository. Documenting each step of data analysis, including raw data, is crucial for reproducibility.
2. Convert all data to a common format so that it can be integrated: Preprocess data to convert to CSV files with a single sheet per file and a single header row. Standardize column headers by mapping to ontology templates. Output as a CSV

file and store on CyVerse under “pre-processed.” Data processing scripts are available on a per dataset basis at https://github.com/UA-SRC-data/data_loaders

3. Visualize and validate data: Loading data into the project MySQL or other database allows for preliminary visualization, the first step in most big data projects. This acts as a validation step that allows us to identify outliers and errors in the data such as incorrect units, mapping errors during step 2, or incorrect datatypes.
4. Run data through the Ontology Data Pipeline (<https://github.com/biocodellc/ontology-data-pipeline>) to convert to graph format. In addition to standardizing the data, the ontology can infer new facts such as hierarchical classification, which enhances searching. More details on the use of ontologies is included in section Standardizing Vulnerability and Resiliency Variables to Ontologies.
5. Output final datasets including standardized versions of datasets for publication as well as complete versions of dataset to use in the visualization portal.

Decision-Making and Standardization Practices—Integrating data from multiple databases requires many decisions regarding which data to include, how to carry forward missing or other special values, and how to harmonize data collected at different spatial scales or time points. It is critical to document these decisions and ensure that documentation accompanies any published datasets.

Managing Non-numerical Values in Numerical Fields: It is common for data sources to include non-numerical values in certain cells where a numerical value is expected. Data from the National Water Quality Monitoring Council includes values like “<0.02” to indicate metal(loid) concentrations below the limit of detection (LOD). Such values will not pass validation and cannot be computed on, so researchers must decide how to use them, which is challenging for data that were collected by a third party. In the Gardenroots dataset, all values below LOD are recorded as LOD/ 2 so that they can be included in analyses (USEPA, 1991; Helsel, 2011). Because our database (step 3 in section Data processing SOP) specifies datatypes (e.g., float, string) for each field, it will automatically find values of this type that need to be addressed.

Variation in Spatial Granularity: Spatial resolution varies among datasets, including both point locations and shape files at the census block, block group, and tract level or county level. We chose the census block group as our preferred spatial resolution because it strikes a good balance between specificity and availability among different data sources, and because it is the resolution of Gardenroots data (see section below on privacy). Furthermore, for some datasets, such as the USEPA’s Environmental Justice Screening tool (EJSCREEN), limited data availability at finer resolution can lead to unacceptable levels of uncertainty (USEPA OECA, 2014). Some datasets are only available at the county level (e.g., USDA data), so any analyses at finer scales must include the uncertainty that comes from using county level data. Data at the point level (e.g., USGS water monitoring data) can be converted to block group using standard code libraries (see https://github.com/UA-SRC-data/data_loaders/tree/master/point2shape), with the recognition that this introduces

uncertainty for that block group. Therefore, processed datasets must include annotations that data were converted from point to shape file. In addition to the spatial resolutions listed above, we are also adding spatial files for different boundary types, to represent, for example, tribal homelands and Primary Care Areas.

Variation in Temporal Granularity: The time of data collection also varied among datasets. Gardenroots data were collected over multiple years (2010–11, 2015–16, 2019), sometimes with multiple data points for the same location. Some federal datasets are available for multiple years, while others are available for a single year only. For those that are available over multiple years (e.g., EJSSCREEN), we chose to use only data from the most recent year. Because our integration uses only datasets from the period of 2010–2020, we make the assumption that they are comparable, but variation in year collected introduces additional uncertainty. Often, when data at a broader temporal resolution are combined with date where an exact date is known, a specific date will be assigned. For example, a data point for 2018 might be assigned a date of January 1, 2018, to be compatible with other data points for which the day of the year is known. This introduces a false sense of precision. When integrating data at different scales (spatial or temporal), the integration must usually happen at the largest scale, even if this means losing information from more precise dataset.

Protecting Privacy of CS Data—When working with CS data such as Gardenroots, it is critical to preserve the privacy of participants. Before data collection, all participants were consented under the University of Arizona Institutional Review Board as an approved project for learning research. Although the UA currently does not see environmental monitoring as a “type” of human research, the name, location, and reported-back environmental monitoring data were deidentified to preserve participant privacy. It is clear that there is an ethical duty to report data back to participants, but once the participant has that data, are there ethical or legal implications? Do they have to disclose when selling their home? Renting? Having family members visit who are considered a member of the sensitive population (i.e., under five, over 65, and/or have a preexisting condition)? Goho (2016) reviewed and explored the potential legal duties of study participants once they have participated in a residential exposure study and have received their personalized data results. It was concluded that there are both ethical and legal implications that researchers and community researchers need to consider, highlighting how data privacy and preservation is critical to CS data science efforts. Based on the above and previous efforts, a solution was reached where community data reported herein was kept to the geographic resolution of the census block when the census block includes at least 10 residences and the census block group otherwise.

Standardizing Vulnerability and Resiliency Variables to Ontologies—Ontologies are standardized terminologies that provide logical (understood by computers) and text (understood by humans) definitions to reduce ambiguity about the meaning of data. A key challenge in integrating such disparate datasets (with variables ranging from metal(loid) concentration in garden soil to median household income to proximity to grocery stores) is the lack of widely adopted ontologies for vulnerability and resiliency factors. Environmental vulnerability terms for chemical exposures have the best existing coverage in ontologies,

due to chemical terms in Chemical Entities of Biological Interest ontology (CheBI, de Matos et al., 2009) and environmental quality terms in the Environment Ontology (ENVO, Buttigieg et al., 2013, 2016). We follow the best practice of submitting new term requests to existing ontologies to fill gaps, but that process can be slow. Therefore, we have created an application ontology, the University of Arizona Superfund Research Project Data Interface Ontology (SRPDIO, <https://github.com/UA-SRC-data/srpdio>) to meet our pressing data integration needs. Superfund Research Project Data Interface Ontology reuses terms from the ENVO, CheBI, the Exposure Ontology (Mattingly et al., 2012), and other ontologies to standardize variable names across datasets. We are working with ENVO curators to move physio-chemical parameters such as metal(loid) concentration or electroconductivity into ENVO, where they can be more broadly reused. For variables that have no ontology (e.g., number of registered voters or proximity to EPA Risk Management Plan Facilities), we are creating terms within the SRPDIO to explicitly define each variable. We plan to work with the larger environmental health community to develop ontologies around social vulnerability and resiliency factors in the future.

RESULTS

Integrated Datasets That Permit New Environmental Health Studies

The Gardenroots CS data were integrated with existing governmental environmental monitoring data to create a more holistic story that includes vulnerability and resiliency data from these rural, medically-underserved communities. We integrated typically siloed/separated datasets including datasets that are intentionally segregated based on who collected the data. The integration of these datasets allowed for the generation of the proposed questions in Table 3 that we anticipate answering (see Figures 3–5 for examples). The vulnerability and resiliency data in Tables 1, 2 are in various stages of processing with the SOP described in section Data processing SOP. Metal(loid) concentration data from Gardenroots, National Water Quality Monitoring Council, and USGS; pollution-related data from US EPA's EJSCREEN; and social data from the U.S. Census Bureau's American Community Survey (ACS) have been preprocessed, and validated using our internal database (step 3 in section Data processing SOP). These datasets are available in an archived release of our GitHub repository (Youens-Clark et al., 2020) in files named "scrutinizer.csv" under their corresponding directories, along with the code that generated them. Because Gardenroots contains multiple datasets, the pre-processed data are instead in a directory named "scrutinizer" with separate files for plant and soil data. Food access data from USDA's Economic Research Service, health data from the National Center for Health Statistics, and NIH's Health Resources and Service Administration's Health Professional Shortage Area data have been downloaded and stored in our shared CyVerse directory, but still require standardization. We do not publish those datasets, as they are available from the original sources, listed in Tables 1, 2. To access data from the Center for Disease Control's Behavioral Risk Factor Surveillance System (BRFSS), we will need to submit a request and complete an application. CDC restricts re-publication of BRFSS data, so although we plan to use them in our integrated data modeling work, we will not be able to publish them. Recognizing the need for privacy, FAIR principles do not require that data be open, but they do require adequate description. Therefore, we will provide full metadata for

any BRFSS data we use. Arizona voter data, available from <https://azsos.gov/precinct-level-results-county-2018-general-election>, will require additional processing for some counties to extract the desired variables, because county-level data are not reported consistently.

Though we have not yet used this methodology extensively, Figures 3–5 are example visualizations generated from the integration of selective datasets at the county level, demonstrating initial and further anticipated results. We recognize that causality cannot be inferred, but these examples show how the database can help inform hypothesis generation and identify counties that are suffering more from selected health outcomes and/or environmental quality challenges. These visualizations were created to align with the questions posed in Table 3 to illustrate the breadth of this methodology. For example, arsenic and chromium (inhalation route only) are recognized as human carcinogens by USEPA, while cadmium and lead are classified as probable carcinogens (e.g. USEPA Integrated Risk Information System, USEPA, 2021). Figure 3 supports hypothesis generation, specifically asking whether arsenic, cadmium, chromium, and/or lead soil concentrations occur in counties with high incidence rates of the most commonly observed cancer types, informing questions 3–4 in Table 3. We see that Mohave county experiences bladder, lung, kidney and pancreatic cancers, but is only impacted by chromium in soil, whereas Yavapai county is impacted by all metal(loid)s except cadmium and the bladder and lung cancer incidence rates are among the top five. Greenlee county has the highest concentrations of cadmium, chromium, and copper (currently not classified as a human carcinogen), however the Arizona Department of Health Services dataset is missing selected cancer incidence rates, which we will gather from another source listed in Table 1.

In addition to cancer, studies have also observed that arsenic exposure is associated with an increased risk of developing a number of diseases, including cardiovascular disease and type II diabetes (Sears and Genuis, 2012; Naujokas et al., 2013). Currently, University of Arizona Superfund researchers are working to determine how chronic exposure to mine wastes that contain arsenic contributes to the development of diabetes. Figure 4 examines the prevalence of diagnosed diabetes and obesity along with major mining activities in Arizona (Niemuth, 2015; Centers for Disease Control and Prevention, 2016; Richardson et al., 2019). Mining and industrial processes are primary sources of arsenic and heavy metal contamination in soil (Lee et al., 2005). Greenlee, Gila, Pinal, Navajo, Graham, La Paz, and Mohave populations have an incidence rate of diabetes and obesity at the medium level, 9–13.9 and 29.1–36.0%, respectively, as well as at least one major mine, informing questions 3–5 in Table 3.

Figure 5 highlights a human capital form of resiliency—the percent of internet subscriptions (dial-up and broadband, cellular data plan, and satellite internet services) in Arizona counties. Internet service is a form of resiliency, indicating potential technical literacy and access to information. The highest percentage of internet service is 35.88% in Yavapai county, followed by Pima, Mohave, and Maricopa counties. This information indicates that researchers, government agencies, and other organizations cannot solely rely on websites for information dissemination, informing question 11 in Table 3.

New Ontology Terms

Metal(loid) Environmental Monitoring Data—The initial draft of the SRPDIO and the code used to generate it are available at <https://github.com/UA-SRC-data/srpdio>, with the first official release in November 2020. A key component of the SRPDIO is the creation of new ontology terms for concentrations of metal(loid)s in environmental materials and plant structures. We use logical definitions for these terms that allow the ontological reasoner to automatically build complex hierarchies of metal(loid) concentrations (Figure 2). The logical definitions (Figures 2B,C) follow an ontology design pattern established in ENVO and the Plant Trait Ontology (TO, Cooper et al., 2018) to define terms for concentrations. “Inheres in” comes from the widely used Relations Ontology (Wg, 2020). It is used to relate a quality (in this case, “concentration of”) and the entity that has that quality (in this case a “plant structure” or a “material entity”). The ontological hierarchies support advanced queries, such as “find all data on any metal in a plant structure” or “find all data on zinc contamination in any material.” These terms and definitions were created in the SRPDIO but will be moved to the Environment Ontology with an upcoming ENVO release.

Sociodemographic Data—Another key component of the SRPDIO is the development of new ontology terms for sociodemographic variables. Currently, there is not a fully developed ontology for sociodemographic data, such as the information collected in the U.S. Census Bureau’s ACS. This was acknowledged as a main concern in Gobel et al. (2016), where interviewed stakeholders reported that current interoperability efforts are biased and limited to the natural sciences. Interviewees were critical that social science standards were absent from discussions, highlighting that any proposed interface and standardization effort would need to be accessible to a wide range of projects and research methodologies. Here, we acknowledge this bias and that the data science efforts have traditionally focused on the natural sciences, entailing observational data, and are not applicable to all forms of knowledge (Gobel et al., 2016). Another issue highlighted by interviewees in the aforementioned Stakeholder Analysis, was the lack of clarity on how to treat data gathered on participants including sociodemographic information and participant evaluations. As highlighted in section Protecting privacy of CS data, we have proposed a solution where community data can be reported while protecting privacy.

DISCUSSION

Solving Environmental Health Challenges With Transdisciplinary Data Science

This data science methods paper demonstrates the integrated framework needed to solve the challenges of interoperability within the environmental health sciences as well as how to integrate CS data. We have developed a methodology to make environmental health CS data FAIR, while also integrating other types of environmental health and social data to enhance discoverability, reuse of data for research translation, and enable hypothesis generation. We are, to the best of our knowledge, among the first to develop ontology terms for:contaminant concentrations in various environmental media, and sociodemographic data. This effort is advancing the field, while also demonstrating how the designed data management system can be applied to other research questions and scenarios. An anticipated result of this

integration effort is that it will help the field determine if and how community-level resiliencies may combat environmental health vulnerabilities.

The complex challenges associated with environmental health and food security are influenced by current and emerging political, social, economic, and environmental contexts. To solve these “wicked” dilemmas (Rittel and Webber, 1973), we need methods to harness the public’s participation in research, conceptualize solutions, and strategize implementations at all levels of the ecological model of health to effectively design interventions (Bronfenbrenner, 1979; Richard et al., 2011). These challenges do not respect disciplinary boundaries. Therefore, transdisciplinary research efforts are needed (*e.g.*, Ramírez-Andreotta et al., 2014; Anderson et al., 2015; Pohl et al., 2017) that follow FAIR principles so that the varying knowledge sources can be interwoven (Anderson et al., 2015; Pohl et al., 2017). Based on the datasets highlighted and integrated in our case study, we do not necessarily need more data, we need integrated data management practices to solve the challenges of interoperability of CS data within the environmental health sciences.

Place-Based Strategies to Mobilize Resiliencies

The data science methods reported here go beyond simply integrating CS environmental vulnerabilities datasets. Citizen and community science efforts can be viewed as place-based strategies to address public health challenges such as health promotion and environmental exposures. To build upon place-based strategies and social processes (*e.g.*, Ness and Powles, 1997; Armstrong, 2000; Teig et al., 2009), we combined CS data with data on the communities’ human, social, and political capital to help inform how rural mining populations can mitigate potential chronic exposures and rebound when their ecosystem has been negatively impacted. For example, to combat natural disasters, Bergstrand et al. (2015) mapped social vulnerability and community resilience to visualize community risks as well as their capacities for recovery and adaptation.

Regarding enabling hypothesis generation, we anticipate that the integrated Garderoots and government datasets will reveal new forms of community resiliency that can be mobilized to support and protect ecosystem services. Community resilience theory has become a key component of national policies across federal agencies because it provides a framework that embraces principles of equity and justice with a focus on building the capacities of populations both to mitigate disasters and to successfully rebound (Norris et al., 2008; Plough et al., 2013). Our methodology builds on this theory and can ultimately help to directly inform decision-making in these communities and identify critical areas for further study (Figures 3–5). Further, we anticipate that an understanding of soil quality from combining Gardenroots and USGS datasets will support provisioning services and inform where local food production efforts should be invested, addressing food deserts that have been highlighted in the USDA data. Alternatively, if local soils are not suited for crops, affected families can be connected to the Supplemental Nutrition Assistance and/or the Woman’s Infant and Children Programs for nutritional assistance. As a second example, in one community we had monthly meetings dedicated to identifying local concerns and priorities. A discussion regarding the need for occupational diversity has been initiated. The community does not want to be solely dependent on a local copper mine for economic

prosperity as they recognize that copper ore is a finite resource. Thus, the community is interested in diversifying the types of jobs available in their community. Understanding current employment rates and labor force status, educational attainment, and the presence of a computing device and internet service/subscription at home, in the context of community educational, recreation and tourism, aesthetic, and cultural heritage values, can illuminate the best investments to make in social and human capital to facilitate occupational and economic diversification while protecting cultural resources.

Determining Community-Level Resiliencies and How They May Combat Environmental Health Vulnerabilities

The resiliency literature has demonstrated that individual, family, social, and environment resources are critical for the successful recovery of a community or cultural system. Resiliency allows a given community to absorb a disturbance. This includes the ability to reorganize to meet the challenges of a change while still retaining the elements that make a community distinct (e.g., Healy, 2006; Fleming and Ledogar, 2008). Unfortunately, past efforts to understand resilience have focused on ecological systems and include socio-ecological systems to a much lesser extent (Bhamra et al., 2011). We anticipate that the lack of consideration of socio-economic systems is due to the absence of available information at different scales and research domains. But we argue that sustained community resiliency heavily relies on the improvement of social factors and this is a missed opportunity. Among these social factors are education, employment, and population well-being (Abramson et al., 2015).

The most important single predictor of health is socioeconomic status (e.g., Singh-Manoux et al., 2018; Kivimäki et al., 2020). Thus, one cannot separate socioeconomic status from environmental health vulnerabilities. However, efforts to improve environmental health need to include a better understanding and mobilization of current community-level resiliencies to help improve the socioeconomic status of the community as a whole. Table 3 illustrates the type of questions related to vulnerability and resiliency that our proposed framework and methodology would enable exploring. The anticipated outcome is illumination of improved and placed-based solutions to environmental health vulnerabilities (see section Place-Based Strategies to Mobilize Resiliencies).

Power and Challenges of Interoperability

We have developed a method to provide a common data model that allows environmental health researchers working at different scales and research domains to exchange data. This method provides the ability to usefully incorporate such data, scaling the impact of any single dataset, be it from a single government, NGO, university, or CS source. We are currently working on an end-user/stakeholder analysis to determine “what works,” “what is missing,” and how to create the interactive data visualization approach that can be used for exploratory analysis and dissemination.

Key tasks for this goal include (Sedlmair et al., 2012): (1) Observations of current end-user/stakeholder’s analytical workflow and data visualization practices to prepare a validated visualization system; (2) Formative evaluation and usability studies of the validated

visualization system with new end-users/stakeholders to ensure the visualizations meet stakeholder's needs and answer their research questions; and (3) Development of a user-friendly web application that will support efforts to streamline data access, visualization, and analyses. In February 2021, we received University of Arizona Institutional Review Board Approval to start this analysis with local, state, federal, and community stakeholders. The new knowledge gained will aid in the creation of similar tools and workflows for use in other scientific contexts.

Modeling population- and factor-wide environmental effects using existing datasets from academia and federal agencies currently faces a number of challenges, including a limited number of samples in environmental datasets, which may prevent researchers from obtaining robust statistical confidence. Our method, which combines multiple data sources, helps to overcome the lack of power in an individual dataset by increasing the number of datasets available. Another key challenge in integrating CS data with public data and making it FAIR is the lack of existing standards and ontologies for environmental health data. The Children's Health Exposure Analysis Resource ontology (Balshaw et al., 2017) provides broad coverage of environmental health indicators but lacks coverage of many important vulnerability and resiliency terms. We encourage environmental health researchers, especially those with knowledge of social and economic factors (which have the poorest ontological coverage) to get involved in community ontology development in order to support future data standardization and integration efforts. Our future work includes contributions to community ontologies such as ENVO and refinement of the SRPDIO.

CONCLUSION

This effort has allowed for the development of a transdisciplinary data management (and eventually visualization) tool that we anticipate, will: (1) Help mitigate the human impacts of exposure to environmental contamination through effective research translation and community engagement driven by stakeholder-engaged research, and (2) Serve as a global resource for human and environmental health issues associated with contamination whether it is from a legacy site (as described in the Gardenroots example) or from a new or ongoing data source. It is expected that the interoperability efforts discussed herein, combined with the future end-user/stakeholder informed and validated data visualizations, will yield new insights into the factors that affect environmental health—both positively and negatively in communities.

ACKNOWLEDGMENTS

The authors are incredibly grateful to all the Gardenroots families and the past and current members of the Integrated Environmental Science and Health Risk Laboratory. In addition, we are thankful to the University of Arizona's Apache, Cochise, Greenlee, and Yavapai Cooperative Extension Offices that served as a Gardenroots home base.

FUNDING

This research was supported by the National Institute of Environmental Health Sciences, NIH through an administrative supplement to the University of Arizona Superfund Research Center grant P42 ES004940 and by the University of Arizona TRIF (Technology and Research Initiative Fund) Center for Environmentally Sustainable Mining.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author/s.

REFERENCES

- Abramson DM, Grattan LM, Mayer B, Colten CE, Arosemena FA, Bedimo-Rung A, et al. (2015). The resilience activation framework: a conceptual model of how access to social resources promotes adaptation and rapid recovery in post-disaster settings. *J. Behav. Health Serv. Res.* 42, 42–57. doi: 10.1007/s11414-014-9410-2 [PubMed: 24870399]
- Adger WN, and Neil Adger W (2006). Vulnerability. *Global Environ. Change* 16, 268–281. doi: 10.1016/j.gloenvcha.2006.02.006
- Allen BL (2018). Strongly participatory science and knowledge justice in an environmentally contested region. *Sci. Technol. Hum. Values* 43, 947–971. doi: 10.1177/0162243918758380
- Anderson BE, Naujokas MF, and Suk WA (2015). Interweaving knowledge resources to address complex environmental health challenges. *Environ. Health Perspect.* 123, 1095–1099. doi: 10.1289/ehp.1409525 [PubMed: 25910282]
- Armstrong D (2000). A survey of community gardens in upstate New York: implications for health promotion and community development. *Health Place* 6, 319–327. doi: 10.1016/S1353-8292(00)10003-7 [PubMed: 11027957]
- Balshaw DM, Collman GW, Gray KA, and Thompson CL (2017). The children’s health exposure analysis resource: enabling research into the environmental influences on children’s health outcomes. *Curr. Opin. Pediatr.* 29, 385–389. doi: 10.1097/MOP.0000000000000491 [PubMed: 28383342]
- Bergstrand K, Mayer B, Brumback B, and Zhang Y (2015). Assessing the relationship between social vulnerability and community resilience to hazards. *Soc. Indic. Res.* 122, 391–409. doi: 10.1007/s11205-014-0698-3 [PubMed: 29276330]
- Bhamra R, Dani S, and Burnard K (2011). Resilience: the concept, a literature review and future directions. *Int. J. Product. Res.* 49, 5375–5393. doi: 10.1080/00207543.2011.563826
- Bronfenbrenner U (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Cambridge, MA; London: Harvard University Press.
- Brown P (1992). Popular epidemiology and toxic waste contamination: lay and professional ways of knowing. *J. Health Soc. Behav.* 33, 267–281. doi: 10.2307/2137356 [PubMed: 1401851]
- Brown P, and Mikkelsen EJ (1997). *No Safe Place: Toxic Waste, Leukemia, and Community Action*. Berkeley, CA: University of California Press. doi: 10.1525/9780520920484
- Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE, and ENVO, Consortium (2013). The environment ontology: contextualising biological and biomedical entities. *J. Biomed. Semantics* 4:43. doi: 10.1186/2041-1480-4-43 [PubMed: 24330602]
- Buttigieg PL, Pafilis E, Lewis SE, Schildhauer MP, Walls RL, and Mungall CJ (2016). The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperability. *J. Biomed. Semantics* 7:57. doi: 10.1186/s13326-016-0097-6 [PubMed: 27664130]
- Centers for Disease Control and Prevention. (2016). Age-Adjusted Prevalence of Obesity and Diagnosed Diabetes Among Adults, by County, United States. Available online at: <https://www.cdc.gov/diabetes/data/center/slides.html> (accessed October 31, 2018).
- Cooper L, Meier A, Laporte M-A, Elser JL, Mungall C, Sinn BT, et al. (2018). The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. *Nucleic Acids Res.* 46, D1168–D1180. doi: 10.1093/nar/gkx1152 [PubMed: 29186578]
- Cooper L, Walls RL, Elser J, Gandolfo MA, Stevenson DW, Smith B, et al. (2013). The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol.* 54:e1. doi: 10.1093/pcp/pcs163 [PubMed: 23220694]
- Cutter SL (1996). Vulnerability to environmental hazards. *Prog. Hum. Geogr.* 20, 529–539. doi: 10.1177/030913259602000407

- Cutter SL, Barnes L, Berry M, Burton C, Evans E, Tate E, et al. (2008). A place-based model for understanding community resilience to natural disasters. *Global Environ. Change* 18, 598–606. doi: 10.1016/j.gloenvcha.2008.07.013
- de Matos P, Alcántara R, Dekker A, Ennis M, Hastings J, Haug K, et al. (2009). Chemical entities of biological interest: an update. *Nucleic Acids Res.* 38(Suppl_1), D249–D254. doi: 10.1093/nar/gkp886 [PubMed: 19854951]
- Dhillon CM (2017). Using citizen science in environmental justice: participation and decision-making in a Southern California waste facility siting conflict. *Local Environ.* 22, 1479–1496; doi: 10.1080/13549839.2017.1360263
- Fleming J, and Ledogar RJ (2008). Resilience, an evolving concept: a review of literature relevant to aboriginal research. *Pimatisiwin.* 6, 7–23. [PubMed: 20963184]
- Garcia AP, Wallerstein N, Hricko A, Marquez JN, Logan A, Nasser EG, et al. (2013). THE (Trade, Health, Environment) impact project: a community-based participatory research environmental justice case study. *Environ. Just.* 6, 17–26. doi: 10.1089/env.2012.0016
- Gobel C, Martin VY, and Ramirez-Andreotta MD (2016). Stakeholder Analysis: International Citizen Science Stakeholder Analysis on Data Interoperability Final Report. Washington, DC: Woodrow Wilson International Center for Scholars. Available online at: https://www.wilsoncenter.org/sites/default/files/stakeholder_analysis_0.pdf
- Goho SA (2016). The legal implications of report back in household exposure studies. *Environ. Health Perspect.* 124, 1662–1670. doi: 10.1289/EHP187 [PubMed: 27153111]
- Gollan J, de Bruyn LL, Reid N, and Wilkie L (2012). Can volunteers collect data that are comparable to professional scientists? A study of variables used in monitoring the outcomes of ecosystem rehabilitation. *Environ. Manage.* 50, 969–978. doi: 10.1007/s00267-012-9924-4 [PubMed: 22875540]
- Gruber T (1993). Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum. Comput. Stud* 43, 907–928. doi: 10.1006/ijhc.1995.1081
- Haklay M (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ. Plann. B Plann. Des.* 37, 682–703. doi: 10.1068/b35097
- Haklay MM (2017). “Volunteered geographic information: quality assurance,” in *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, eds Richardson R, Castree N, Goodchild MF, Kobayashi A, Liu W, and Marston RA (Hoboken, NJ: Wiley-Blackwell), 1–6. doi: 10.1002/9781118786352.wbieg0602
- Healy S (2006). “Cultural resilience, identity and the restructuring of political power in Bolivia,” Paper Submitted for the 11th Biennial Conference of the International Association for the Study of Common Property (Bali). Available online at: http://www.indiana.edu/iascp/bali/papers/Healey_susan.pdf
- Hecker S, Haklay M, Bowser A, Makuch Z, and Vogel J (2018). *Citizen Science: Innovation in Open Science, Society and Policy*. London: UCL Press. doi: 10.2307/j.ctv550cf2
- Helsel DR (2011). *Statistics for Censored Environmental Data Using Minitab and R*. Hoboken, NJ: John Wiley and Sons. doi: 10.1002/9781118162729
- Kivimäki M, Batty GD, Pentti J, Shipley MJ, Sipilä PN, Nyberg ST, et al. (2020). Association between socioeconomic status and the development of mental and physical health conditions in adulthood: a multi-cohort study. *Lancet. Public Health* 5, E140–E149. doi: 10.1016/S2468-2667(19)30248-8 [PubMed: 32007134]
- Landrigan PJ, Fuller R, Acosta NJR, Adeyi O, Arnold R, Basu N, et al. (2018). The Lancet Commission on pollution and health. *Lancet* 391, 462–512. doi: 10.1016/S0140-6736(17)32345-0 [PubMed: 29056410]
- Lee J-S, Chon H-T, and Kim K-W (2005). Human risk assessment of As, Cd, Cu and Zn in the abandoned metal mine site. *Environ. Geochem. Health* 27, 185–191. doi: 10.1007/s10653-005-0131-6 [PubMed: 16003586]
- Manjón I, Ramírez-Andreotta MD, Sáez AE, Root RA, Hild J, Janes MK, et al. (2020). Ingestion and inhalation of metal(loid)s through preschool gardening: an exposure and risk assessment in

- legacy mining communities. *Sci. Total Environ.* 718:134639. doi: 10.1016/j.scitotenv.2019.134639 [PubMed: 31843310]
- Mattingly CJ, McKone TE, Callahan MA, Blake JA, and Hubal EAC (2012). Providing the missing link: the exposure science ontology ExO. *Environ. Sci. Technol.* 46, 3046–3053. doi: 10.1021/es2033857 [PubMed: 22324457]
- Merchant N, Lyons E, Goff S, Vaughn M, Ware D, Micklos D, et al. (2016). The iPlant collaborative: cyberinfrastructure for enabling data to discovery for the life sciences. *PLoS Biol.* 14:e1002342. doi: 10.1371/journal.pbio.1002342 [PubMed: 26752627]
- Morello-Frosch R, Brody JG, Brown P, Altman RG, Rudel RA, and Pérez C (2009). Toxic ignorance and right-to-know in biomonitoring results communication: a survey of scientists and study participants. *Environ. Health* 8:6. doi: 10.1186/1476-069X-8-6 [PubMed: 19250551]
- Nagy C, Bardwell K, Rockwell RF, Christie R, and Weckel M (2012). Validation of a citizen science-based model of site occupancy for eastern screech owls with systematic data in suburban New York and Connecticut. *Northeast. Natur.* 19, 143–158. doi: 10.1656/045.019.s611
- Naujokas MF, Anderson B, Ahsan H, Aposhian HV, Graziano JH, Thompson C, et al. (2013). The broad scope of health effects from chronic arsenic exposure: update on a worldwide public health problem. *Environ. Health Perspect.* 121, 295–302. doi: 10.1289/ehp.1205875 [PubMed: 23458756]
- Ness AR, and Powles JW (1997). Fruit and vegetables, and cardiovascular disease: a review. *Int. J. Epidemiol.* 26, 1–13. doi: 10.1093/ije/26.1.1 [PubMed: 9126498]
- Niemuth N (2015). Arizona Major Mines Map - 2015. AZGS Document Repository [map]. Arizona Geological Survey. Available online at: http://repository.azgs.az.gov/uri_gin/azgs/dlio/1670
- Norris FH, Stevens SP, Pfefferbaum B, Wyche KF, and Pfefferbaum RL (2008). Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. *Am. J. Commun. Psychol.* 41, 127–150. doi: 10.1007/s10464-007-9156-6
- Ottinger G (2010). Buckets of resistance: standards and the effectiveness of citizen science. *Sci. Technol. Hum. Values* 35, 244–270. doi: 10.1177/0162243909337121
- Pandya RE (2012). A framework for engaging diverse communities in citizen science in the US. *Front. Ecol. Environ.* 10, 314–317. doi: 10.1890/120007
- Plough A, Fielding JE, Chandra A, Williams M, Eisenman D, Wells KB, et al. (2013). Building community disaster resilience: perspectives from a large urban county department of public health. *Am. J. Public Health* 103, 1190–1197. doi: 10.2105/AJPH.2013.301268 [PubMed: 23678937]
- Pocock MJO, Tweddle JC, Savage J, Robinson LD, and Roy HE (2017). The diversity and evolution of ecological and environmental citizen science. *PLoS ONE* 12:e0172579. doi: 10.1371/journal.pone.0172579 [PubMed: 28369087]
- Pohl C, Truffer B, and Hirsch-Hadorn G (2017). “Addressing wicked problems through transdisciplinary research,” in *The Oxford Handbook of Interdisciplinarity*, Frodeman R (Oxford: Oxford University Press), 319–331. doi: 10.1093/oxfordhb/9780198733522.013.26
- Ramírez-Andreotta MD, Brusseau ML, Artiola J, Maier RM, and Gandolfi AJ (2015). Building a co-created citizen science program with gardeners neighboring a superfund site: the Gardenroots case study. *Int. Public Health J.* 7:13. [PubMed: 25954473]
- Ramírez-Andreotta MD, Brusseau ML, Artiola JF, and Maier RM (2013a). A greenhouse and field-based study to determine the accumulation of arsenic in common homegrown vegetables grown in mining-affected soils. *Sci. Total Environ.* 443, 299–306. doi: 10.1016/j.scitotenv.2012.10.095 [PubMed: 23201696]
- Ramírez-Andreotta MD, Brusseau ML, Artiola JF, Maier RM, and Gandolfi AJ (2014). Environmental research translation: enhancing interactions with communities at contaminated sites. *Sci. Total Environ.* 497–498, 651–664. doi: 10.1016/j.scitotenv.2014.08.021
- Ramírez-Andreotta MD, Brusseau ML, Beamer P, and Maier RM (2013b). Home gardening near a mining site in an arsenic-endemic region of Arizona: assessing arsenic exposure dose and risk via ingestion of home garden vegetables, soils, and water. *Sci. Total Environ.* 454–455, 373–382. doi: 10.1016/j.scitotenv.2013.02.063
- Ramírez-Andreotta MD, Tapper A, Clough D, Carrera JS, and Sandhaus S (2019). Understanding the intrinsic and extrinsic motivations associated with community gardening to improve environmental

public health prevention and intervention. *Int. J. Environ. Res. Public Health* 16:494. doi: 10.3390/ijerph16030494

Richard L, Gauvin L, and Raine K (2011). Ecological models revisited: their uses and evolution in health promotion over two decades. *Annu. Rev. Public Health* 32, 307–326. doi: 10.1146/annurev-publhealth-031210-101141 [PubMed: 21219155]

Richardson CA, Swartzbaugh L, Evans T, and Conway FM (2019). Plate 1: Active Mines of Arizona. Directory of Active Mines in Arizona: FY 2019 [map]. Tucson, AZ: Arizona Geological Survey.

Rittel HWJ, and Webber MM (1973). Dilemmas in a general theory of planning. *Policy Sci.* 4, 155–169. doi: 10.1007/BF01405730

Sandhaus S, Kaufmann D, and Ramirez-Andreotta M (2019). Public participation, trust and data sharing: gardens as hubs for citizen science and environmental health literacy efforts. *Int. J. Sci. Educ. B. Commun. Public Engagem.* 9, 54–71. doi: 10.1080/21548455.2018.1542752 [PubMed: 31485378]

Sears ME, and Genuis SJ (2012). Environmental determinants of chronic disease and medical approaches: recognition, avoidance, supportive therapy, and detoxification. *J. Environ. Public Health* 2012:356798. doi: 10.1155/2012/356798 [PubMed: 22315626]

Sedlmair M, Meyer M, and Munzner T (2012). Design study methodology: reflections from the trenches and the stacks. *IEEE Trans. Visual. Comp. Graph.* 18, 2431–2440. doi: 10.1109/TVCG.2012.213

Shirk JL, Ballard HL, Wilderman CC, Phillips T, Wiggins A, Jordan R, et al. (2012). Public participation in scientific research: a framework for deliberate design. *Ecol. Soc.* 17:29. doi: 10.5751/ES-04705-170229

Singh-Manoux A, Fayosse A, Sabia S, Tabak A, Shipley M, Dugravot A, et al. (2018). Clinical, socioeconomic, and behavioural factors at age 50 years and risk of cardiometabolic multimorbidity and mortality: a cohort study. *PLoS Med.* 15:e1002571. doi: 10.1371/journal.pmed.1002571 [PubMed: 29782486]

Teig E, Amulya J, Bardwell L, Buchenau M, Marshall JA, and Litt JS (2009). Collective efficacy in Denver, Colorado: strengthening neighborhoods and health through community gardens. *Health Place* 15, 1115–1122. doi: 10.1016/j.healthplace.2009.06.003 [PubMed: 19577947]

Tregidgo DJ, West SE, and Ashmore MR (2013). Can citizen science produce good science? Testing the OPAL Air Survey methodology, using lichens as indicators of nitrogenous pollution. *Environ. Pollut.* 182, 448–451. doi: 10.1016/j.envpol.2013.03.034 [PubMed: 23631940]

USEPA (1991). Chemical Concentration Data Near the Detection Limit. Region III, Office of Superfund Hazardous Waste Management, EPA/903/8–9/001.

USEPA (2013). Iron King Mine - Humboldt Smelter Site Profile. Available online at: <https://cumulis.epa.gov/supercpad/cursites/csinfo.cfm?id=0905049>

USEPA (2020). Citizen Science for Environmental Protection. <https://www.epa.gov/citizen-science>.

USEPA (2021). Integrated Risk Information System (IRIS). Washington, DC: Office of Research and Development, National Center for Environmental Assessment. Available online at: <http://www.epa.gov/iris/> (accessed April 27, 2021).

USEPA and OECA (2014). Limitations and Caveats in Using EJSscreen. Available online at: <https://www.epa.gov/ejscreen/limitations-and-caveats-using-ejscreen>

Wg OT (2020). Relation Ontology Available online at: <http://www.obofoundry.org/ontology/ro.html> (Retrieved October 21, 2020).

Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18 [PubMed: 26978244]

Youens-Clark K, Walls RL, Blumberg K, Ramirez-Andreotta MD, and Kaufmann D (2020). UA-SRC-data/data_loaders: v1.0 (Version v1.0). Zenodo. Available online at: https://zenodo.org/record/4116472#.X5H_N0JKjUJ

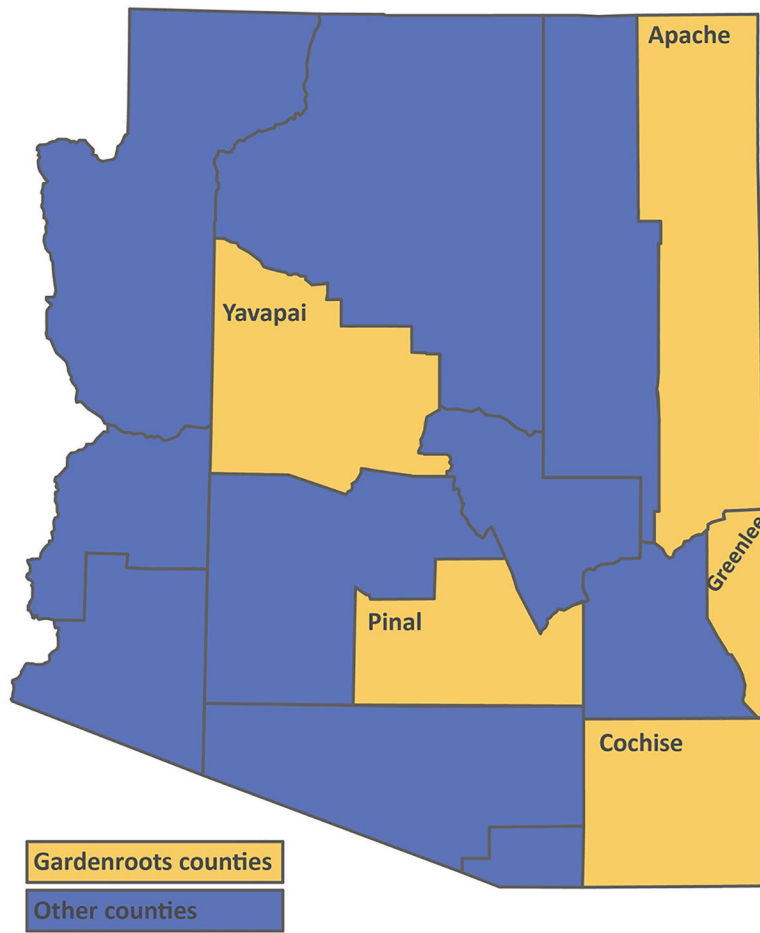


FIGURE 1 |
Map of participating Gardenroots communities in Arizona.

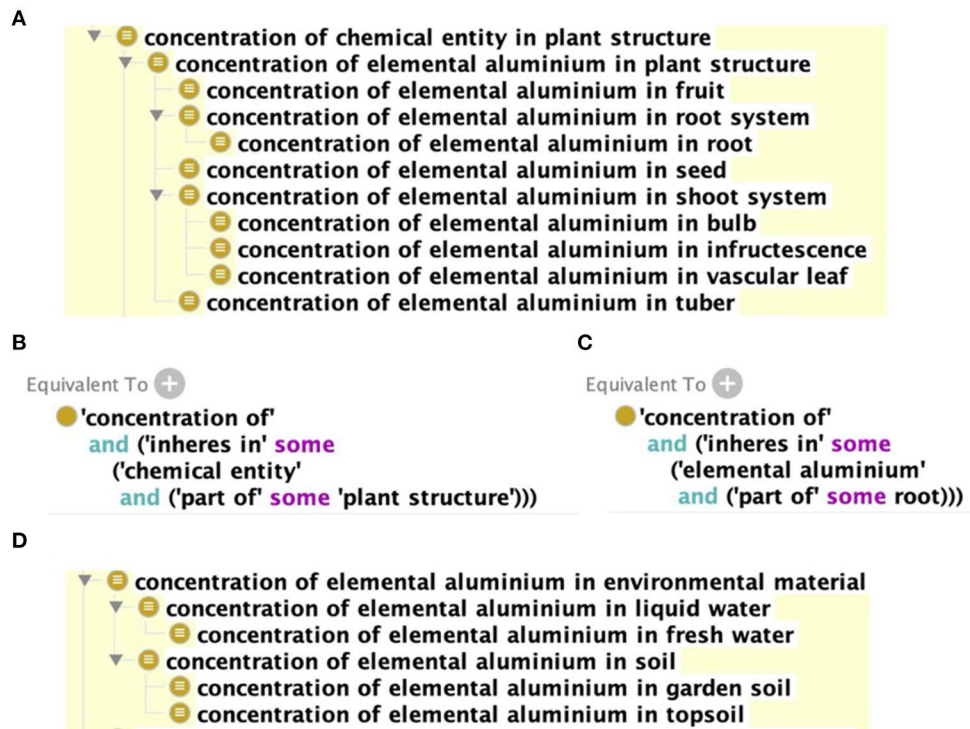


FIGURE 2 |

(A) “the hierarchy of terms for the “concentration of aluminium in plant structure.” Plant structure terms are imported from the Plant Ontology (Cooper et al., 2013) and chemical terms are imported from ChEBI. (B) The logical definition of “concentration of chemical entities in plant structure.” (C) The logical definition of “concentration of elemental aluminium in root.” An ontological reasoner uses these logical definitions to infer the hierarchy shown in (A). (D) The hierarchy for “concentration of elemental aluminium in environmental material” is generated similarly to the hierarchy for concentrations in plant structures. Note that ChEBI is an international ontology that uses the British spelling “aluminium” shown in the figure, but our search engine includes the American spelling “aluminum”.

Incidence Rates for the Top Five Cancers in Arizona Counties

Source: ADHS Environmental Public Health Tracking Data (2012-2016)

Soil Concentrations for the Top Five Contaminants in Arizona Counties

Source: USGS Soil Characteristics (2013)

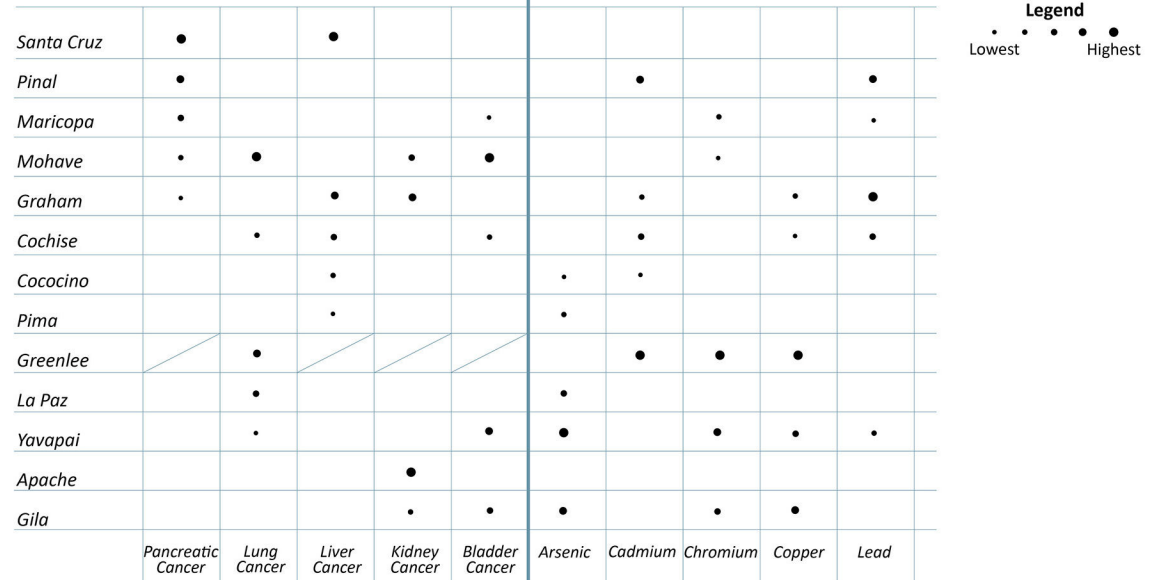


FIGURE 3 |
A visuazation generated from selective datasets to qualitatively describe cáncer incidence rates and soil contaminant concentrations.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

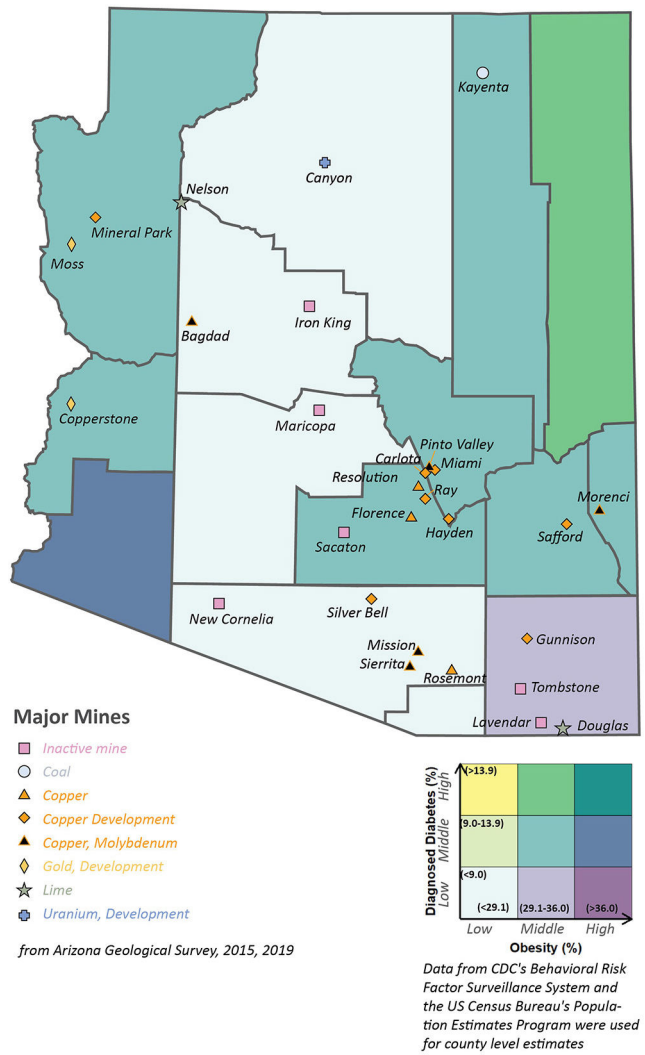


FIGURE 4 | Major mines and prevalence of diagnosed diabetes and obesity in the state of Arizona.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

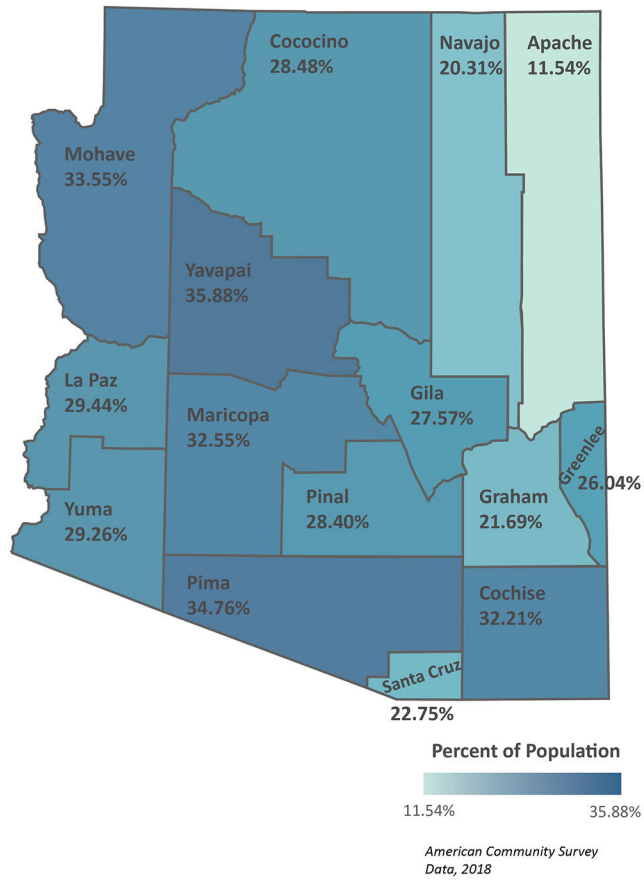


FIGURE 5 |
The percent of county populations with internet access

TABLE 1 |

Vulnerability datasets.

Data description	Dataset source	Year	Variable categories (n = number of variables in the category)
Social attributes	USEPA's EJSCREEN	2019	Linguistically isolated ($n = 3$) Low income ($n = 3$) Minority population ($n = 3$) Less than highschool education ($n = 3$) Under age 5 ($n = 3$) Over age 64 ($n = 3$) Poverty status ($n = 11$)
Quality of environment	American Community Survey <i>Gardenroots</i> Data	2019 2011–2019	Concentrations of metal(loid)s in water ($n = 11$) Concentrations of metal(loid)s in soil ($n = 11$) Concentrations of metal(loid)s in plants ($n = 11$) Concentrations of metal(loid)s in dust ($n = 11$) Proximity to sources of pollution ($n = 30$) Air pollution ($n = 21$) Ozone level in air ($n = 3$) PM 2.5 in air ($n = 3$) Lead paint indicator ($n = 3$) Water quality ($n = 22$) Soil characteristics ($n = 42$) Arsenic in community water systems ($n = 1$) Proximity of population and schools to highways
Quality of health	USEPA's EJSCREEN National Water Quality Monitoring Council U.S. Geological Survey Arizona Department of Health Services (ADHS) Environmental Health Public Tracking Behavioral Risk Factor Surveillance ADHS Environmental Health Public Tracking	2019 2018 2013 2019 2019 2018 2016	Diabetes ($n = 1$) Cancer ($n = 2$) Asthma ($n = 2$) Incidence of cancer ($n = 14$) Hospitalizations for asthma ($n = 3$) Emergency department visits for asthma ($n = 3$) Prevalence of obesity or severe obesity among adults ($n = 1$) State food insecurity ($n = 6$)
Access to food	National Center for Health Statistics U.S. Department of Agriculture's Economic Research Service (USDA ERS)	2020 2020	

The most recent data available was used, with the exception of Gardenroots data, that spans from 2011 to 2020. Since the majority of the datasets had a large amount of measured variables, variables are grouped in categories in this table. The n value represents the number of variables in each category.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 2 |

Resiliency datasets.

Data description	Dataset source	Year	Variable categories (n = number of variables in the category)
Economic capital	American Community Survey	2019	Mortgage (<i>n</i> = 3) Employment (<i>n</i> = 782) Labor force status (<i>n</i> = 279)
Human capital	American Community Survey	2019	Education attainment (<i>n</i> = 27) Healthcare coverage (<i>n</i> = 25) Presence of computing device (<i>n</i> = 11) Internet service/subscription (<i>n</i> = 57) Access to women's infants and children program (<i>n</i> = 15) Access to supplemental nutrition assistance program (<i>n</i> = 20)
Political capital	Arizona Secretary of State	2018	Registered voters (<i>n</i> = 1) Ballots casted (<i>n</i> = 1) Access to polling places (<i>n</i> = 1)
Social capital	Gardenroots Data	2013–2019	Number of Arizona communities participating in <i>Gardenroots</i> (<i>n</i> = 5)
	USDA ERS	2020	Proximity to grocery stores (<i>n</i> = 41) Store availability (<i>n</i> = 35) Food assistance (<i>n</i> = 59) Local foods (<i>n</i> = 96) Restaurants (<i>n</i> = 15)
	Human Resources and Service Administration-Health Professional Shortage Area	2019	Federally qualified health center (<i>n</i> = 1) Indian, tribal, and urban Indian organizations (<i>n</i> = 1) State mental hospital (<i>n</i> = 1)
	ADHS Environmental Health Public Tracking	2019	Rural health clinic (<i>n</i> = 1) Access to parks and elementary schools (<i>n</i> = 1)
		2013	Land use (<i>n</i> = 2)

The most recent data available was used, with the exception of Gardenroots data that spans from 2011 to 2020. Since the majority of the datasets had a large number of measured variables, the variables are grouped in categories. The n value represents the number of variables in each category.

TABLE 3 |

Questions to ask of the vulnerability and resiliency dataset to achieve environmental justice in communities neighboring active and legacy mining activities.

Questions	Dataset used
1. What is/are the major: a. Vulnerability(ies) b. Resiliency(ies)	All datasets in Tables 1, 2
2. Are we (all stakeholders) addressing them? a. If not, how can we?	N/A
3. Are mining communities disproportionately exposed to Arsenic? a. If so, what is/are the major arsenic contributor(s) to daily dose of arsenic?	<ul style="list-style-type: none"> • <i>Gardenroots Data</i> • USEPA's EJSCREEN • National Water Quality Monitoring Council • U.S. Geological Survey
4. Are mining communities suffering/experiencing cancer/diabetes/obesity/asthma disproportionately? a. Why? b. Is it due to rural health disparities? c. Access to nutritional foods and public health programming?	<ul style="list-style-type: none"> • ADHS Environmental Health Public Tracking • Behavioral Risk Factor Surveillance • National Health and Nutrition Examination Study • USDA ERS
5. Are mining communities with elevated arsenic concentrations suffering/experiencing cancer/diabetes/obesity/asthma disproportionately? a. If so, what is/are the major arsenic contributor(s) to daily dose of arsenic?	<ul style="list-style-type: none"> • Gardenroots Data • USEPA's EJSCREEN • National Water Quality Monitoring Council • U.S. Geological Survey • ADHS Environmental Health Public Tracking • Behavioral Risk Factor Surveillance • National Health and Nutrition Examination Study • USDA ERS
6. Can we assign an index value (Bergstrand et al., 2015) ?	All datasets in Tables 1, 2
7. Once we combine the vulnerability and resiliencies, can we rate and compare communities?	N/A
8. How can we leverage the resiliencies to address the vulnerabilities?	All datasets in Table 2
9. When considering ecosystem functions, what function(s) are in deficit/not working? Which functions are working? a. Provisioning b. Regulating c. Cultural d. Supporting	All datasets in Tables 1, 2
10. When considering sustainability practices, what needs to occur: a. Economically—new job opportunities? b. Socially c. Environmentally	All datasets in Tables 1, 2
11. How can we successfully communicate with these communities at the local community and government level?	<ul style="list-style-type: none"> • American Community Survey