THE ROUTLEDGE HANDBOOK OF THE POLITICAL ECONOMY OF THE ENVIRONMENT
Edited by Éloi Laurent and Klara Zwickl

22

BUILDING ON THE RIGHT TO KNOW

Data interlinkage and information intermediation for environmental and corporate regulation

Richard Puchalsky, Michael Ash and James K. Boyce

Introduction

Environmental injustice is systematic exclusion from environmental goods and exposure to environmental "bads" based on social difference, often in ascribed categories such as caste or, especially in the US, race and ethnicity. One of the fundamental tasks in researching environmental justice is to link pollution sources to people affected, which then enables systematic analysis of the social correlates and determinants of exposure.

Arguments over what constitutes proximity between polluting facilities and the people potentially exposed haunted the early years of environmental justice research. The method of unit-hazard coincidence – which finds exposure only for the population residing in the same geographic unit as the polluting facility, be it small area geography of the US Census (tracts or block groups, corresponding more or less to neighborhoods), postal codes, counties, states, or fixed-radius circular buffers has substantive weaknesses (Mohai and Saha 2003), as the shared geography unit may be either too large or too small to make a substantive determination regarding exposure.

Innovative US Environmental Protection Agency (EPA) databases combine information about the type, amount, and geographic origin of pollution with modeled transport to receptor locations that can be associated with US Census data. Linkages between databases are technically complex, often involving imperfect matching methods across linkages of facility ownership, function and scale, and geography and proximity. A fundamental problem is that some applications of data were not envisioned by the entities that produce the underlying databases. For instance, linking individual polluting facilities to the parent companies that own them can relate causes of environmental injustice to corporate policies, but EPA typically does not analyze pollution by parent company even though this information is collected as part of several EPA databases.

In this chapter, we describe broadening uses of two emissions databases from the US EPA that have been important for empirical environmental justice research: the Toxic Release Inventory (TRI), which provides annual, chemical-specific, quantified amounts of air, water, and land pollution from industrial facilities, and the Greenhouse Gas Reporting Program (GHGRP)

DOI: 10.4324/9780367814533-24

database, which provides CO_2-equivalent emissions of greenhouse gases from large, fixed sites. The discussion is technical and detailed; it is intended to function as a guide for practitioners to carry out analysis of environmental justice and corporate environmental justice performance using databases and tools. Availability of data may vary across countries and context. For example, the pollutants listed and the industries covered in pollutant release and transfer registers varies across countries, although the European Pollutant Release and Transfer Register (E-PRTR) harmonizes consistent data collection across Europe. The public availability of data on corporate ownership of pollution sources and on economic and employment characteristics of facilities varies substantially. Census and household survey data vary substantially in terms of geographic detail – many countries do not report data at the neighborhood level – and of the social and economic variables reported. France, notably, does not collect data on national origin for descendants of immigrants or on ethnicity.

The TRI additionally supports yearly estimates of the potential chronic human health risk from each facility by running the underlying data through a value-added EPA model, RSEI (Risk-Screening Environmental Indicators), which produces comparative estimates of chronic human health risk at finely distinguished geographic locations of receptors. RSEI was initially developed at the US EPA to assist in setting priorities for investigation and enforcement among TRI reporting facilities. Hence, its initial structure was designed to generate a univariate comparative risk score for each facility, reflecting the potential chronic human health risk from the facility based on the quantity, toxicity, fate and transport, and population exposure of its emissions. The potential of the high-resolution geographic model of fate and transport was soon recognized by its developers as a potential tool for assessing differential subpopulation risk as well as total population risk (Bouwes et al. 2003).

The GHGRP is designed for reporting greenhouse gas emissions, which are themselves a global problem with limited health impact near the point of release. But greenhouse gas emissions are often accompanied by co-pollutants with substantial local and regional effects, such as particulate matter, nitrogen oxides, and sulfur dioxide, and hence greenhouse gases can be used as a proxy for the local effects of fossil fuel combustion and for environmental justice profiles.

We have linked the TRI/RSEI and GHGRP with each other and with sources of information. The data enable us to rank corporations based on airborne potential chronic human health risk generated by all of their TRI-reporting facilities. The rankings, which are produced on a regular basis from annually updated data from the US EPA, are used by socially responsible investors, corporate environmental managers, regulators, and activists interested in assessing environmental performance at the level of the corporation as well as individual facilities. These data are published on a maintained website, toxic100.org, and are also made available for download to researchers. Using the methodology developed in Ash and Boyce (2011), we characterize the distribution across vulnerable environmental justice communities of chronic human health effects caused by toxic air pollution, both for individual facilities and for parent companies that own them. We additionally characterize the distribution across vulnerable environmental justice communities of co-pollutants of fossil fuel combustion by facilities and the companies that own them.

The data also permit estimation of pollution levels at key receptor sites, including schools, neighborhoods, and other socially vulnerable locations, which enables community, regulatory, and corporate response and permits generalizable research and hypothesis testing about the underlying political and economic processes that lead to differential exposure. Interlinkage of parent company information on emissions of toxics and greenhouse gases to company data on the receipt of public subsidies and the assessment of fines and penalties for environmental, labor, financial, and other infractions enables a more comprehensive characterization of corporate environmental and social governance. Interlinkage of pollutant risk scores with facility-level administrative data on employment from the Equal Employment Opportunity Commission has permitted generalizable research on the often-posited trade-off between jobs and the environment, in particular concerning the employment of minorities in relation to their disproportionate exposure to industrial pollution, and could in principle be applied to holding facilities and companies responsible for their broad social and environmental impact.

Developing interlinkages presents substantial technical challenges, from the integrative assessment model that describes spatial exposure to pollutant releases to selecting and joining spatial, organizational, and socioeconomic data to characterize the landscape of risk. We describe the development of the underlying pollutant release and transfer register data and the process and challenges of linkage. We then briefly survey the results of the studies enabled by the linkages.

We first introduce the toxic air pollutant data used with sections titled "The Toxics Release Inventory: The World's First PRTR" and "Risk Screening Environmental Indicators: An Integrative Assessment Model to Estimate Human Health Risk from Industrial Air Pollution," and then go on to describe how to improve these data with sections titled "Parent Assignment: Corporate Research to Assign TRI Facilities to Ultimate Owners" and "Environmental Justice Ratios: Measuring Corporate Environmental Justice Performance." Subsequent sections introduce the greenhouse gas reporting data used with "Linking Local and Global Pollutants: The Greenhouse Gas Reporting Program" and describe how to improve it with "Extending EJ to Greenhouse Gas Emitters," "Assigning Responsibility for Greenhouse Gas Emissions," and "GHGRP EJ Ratio Analysis for Parent Companies." The section "Lessons on Linkages" describes how to link US EPA databases together at the facility level. Finally, the section "Applied and Social Scientific Research" describes previous research that has been done using these data and "Public Intermediation for Policy Impact" describes previous efforts to provide free and useful public access to the data.

The toxics release inventory: the world's first PRTR

Created by the Emergency Preparedness and Community Right-to-Know Act of 1986 (EPCRA), the Toxics Release Inventory became the world's first national pollutant release and transfer register (PRTR), an innovation that has since been adopted in many industrialized countries. Annual TRI data collection began in 1987, and data from 1988 forward are considered to be of high quality. The TRI was an important advance for researchers because it provides yearly reports of the total mass of toxic chemicals released, broken down by facility, chemical, and medium (air, land, and water) – data that were previously impossible to obtain.

Right-to-know regulation envisions policy or market-based changes resulting from better public information, and right-to-know regulation mandates disclosure of information. Key examples of public information mandates include the pollution data we discuss here and other environmental data such as residential water quality, energy efficiency to guide the purchase of cars and consumer durables, disclosure of the risks inherent in financial assets and loans for investors and consumers, lending performance by banks especially in the domain of racial equity, school performance on standardized tests, and health care provider performance on a range of indicators. Right-to-know regulation often emerges as a compromise between public demands for more concrete regulation and industry resistance to outright regulation. Fung et al. (2007) survey and analyze many domains of regulation by disclosure.

The conversion of right-to-know data into concrete fulfillment of the right to a clean and safe environment requires not only that stakeholders have access to the information, but also, critically, that they have the ability to interpret the information and the capacity and incentive to respond to it (Hersh 2006).

TRI requires large industrial facilities in the United States to report on an annual basis what quantity of each of roughly 600 different listed toxic chemicals or chemical categories the facility released into the environment over the course of the past year. A facility must report to the TRI if it operates in a TRI-covered industrial sector based on its North American Industrial Classification System (NAICS, which replaced the Standard Industrial Classification system), employs at least 10 full-time equivalent employees, and manufactures, processes, or otherwise uses quantities of the chemical in excess of published, chemical-specific thresholds. Most facilities engaged in manufacturing, electrical energy generation from fossil fuels and biomass, coal and metal mining, wholesaling of petroleum products and other chemicals, and hazardous waste storage and disposal must report. Noteworthy exclusions include fracking and oil extraction, electrical energy generation from natural gas combustion, mobile sources, ports, and airports.

The facility, rather than the company, is the reporting unit for TRI and more generally is the object of permitting and enforcement by the US EPA. Each facility submits annual Form R reports to the TRI Central Data Exchange signed by the facility's certifying official. The Form R reports the quantity of each toxic chemical released and the release media (i.e., whether the chemical was released to air, land, or water or transferred offsite). Air media include fugitive releases, stack releases, incineration-based releases, and off-gassing of effluents from publicly operated treatment works. While the term "fugitive release" may evoke industrial accidents and spills, which are subject to reporting requirements, most of the data reported to the TRI (including the fugitive-release category) involves business-as-usual releases. Quantities are reported on an annual basis with no indication of the timing of releases. In their reporting, facilities may use direct measurement of the mass of inputs and outputs or alternative methods, including estimates based on engineering specifications for particular industrial processes.

The TRI data are mandatory and standardized across industries and states, but they are self-reported. While penalties for failing to report and for misreporting are, in principle, high, the occasional TRI enforcement that does occur is generally for nonreporting, and there are indications of systematic underreporting.[1]

The Form R reporting instrument includes a parent company field for facilities to report their corporate ownership, which EPA has attempted to standardize. Remarkably, the US does not maintain a single unique public identifier of ultimate corporate parents. The US EPA uses the Dun and Bradstreet DUNS Number, a private and proprietary unique nine-digit identifier for businesses. The frequency of reorganizations of corporate ownership via mergers, acquisitions, and divestments often leaves parent company information out of date, and the quality of this information, in terms of accurate assignment of facilities to final parent companies, is poor.

Early analysis of TRI data generally consisted of adding up the pounds of releases across chemicals and media for the entities under consideration, be they geographical areas, industrial sectors, or individual facilities. This was unsatisfactory for many purposes because some TRI chemicals are far more hazardous to human health than others. Adding them together by pounds meant that the total often was dominated by lower-risk chemicals released in large quantities, rather than identifying high-risk, low-volume chemicals. EPA's RSEI model was developed to address this and similar issues. RSEI uses a peer-reviewed system of toxicity weights that express how dangerous each chemical is on a per-pound basis; the toxicity weights make it easier to understand the importance of obscurely named chemicals for actual human health risk.

Risk Screening Environmental Indicators: an integrative assessment model to estimate human health risk from industrial air pollution

The US EPA's Risk Screening Environmental Indicators model uses various sources of toxicological information to weigh TRI-listed chemicals for both cancer and non-cancer human health effects, which are put into the same scoring system. RSEI then runs a fate-and-transport model for each TRI facility using weather patterns, velocity and altitude of release, and physico-chemical properties of the released chemical to estimate where air pollution from the facility goes. The estimates are computed for each 810 m by 810 m cell within 50 km of the releasing facility. (RSEI uses a different exposure model for surface water pollution.) Finally, RSEI multiplies the amount of pollution at each receptor location by the number of people residing in that location. The resulting estimate constitutes a comparative risk score that can be added up over any subset of releases and that takes into account the release quantity, chemical toxicity, fate and transport, and the size of the exposed population. A typical EPA use of RSEI is to add up the risk score for each TRI release from a facility over the entire area for which population exposure is computed to establish a risk score for the facility as a whole.

As part of the production of RSEI, estimates of air exposure to each chemical from each facility are made for each grid cell within 50 km of each releasing facility. This allows coverage of the entire US with a consistent closely spaced, high-resolution geographic grid that can be associated with US Census geographic areas. Because the data are broken out by both chemical identity and the individual facility releasing the pollution, the dataset allows for many kinds of analysis. The data providing the toxicity-weighted concentrations for every 810 m by 810 m receptor site, by source facility and chemical, are referred to as the RSEI Geographic Microdata (RSEI-GM). These data are free and publicly available.

Production and use of the RSEI-GM data present several challenges. First, the production is both data intensive, requiring facility-specific information that may not be included in typical PRTR data collection, and computationally intensive, requiring the estimation of a concentration based on a plume model at roughly 12,000 sites for each release from each facility (for roughly 100,000 air releases in 2017).

Second, the RSEI-GM data are very large, requiring substantial computing facilities simply to maintain and access the data. For example, the 2017 data included roughly 1.1 billion data points, each characterizing the effect of one release on one grid cell.

Third, the RSEI-GM grid cells are labelled with an RSEI-specific X-Y coordinate system, which requires some geographic sophistication to use. These X-Y locations can be converted to or from lat-long coordinates, and the US EPA publishes a full crosswalk between X-Y cells and US Census Bureau blocks, the finest geographic unit for the census. With the crosswalk, RSEI concentrations can be compared with or aggregated to US Census American Community Survey five-year data, which contain demographic information suitable for environmental justice research, at the census block group or tract level.

A unique feature of the RSEI model is its tight coupling of source and receptor in the analysis of risk from industrial toxic pollution. Datasets such as the US EPA's National Air Toxics Assessment (NATA) have some advantages over RSEI with respect to the wider range of included pollution sources, including mobile and so-called area sources in addition to the industrial point sources included in TRI. NATA also reports airborne risk from these activities on a high-resolution geographic basis. But the inclusion of multiple sources comes at the cost of losing the association between specific sources and community receptors. As RSEI fully models each toxic release from each releasing facility and maintains release-specific exposure data, it is possible to attach the community burden at the receptor location to the source facility.

This association enables two scorings of facilities: one based on the total potential chronic human health risk from the facility, called the RSEI score; and another based on the potential chronic human health risk for populations and subpopulations of interest, for example, the Hispanic RSEI score indicating the total potential chronic human risk from the facility for the Hispanic, or Latino, population. The subpopulation-specific scores for the facility sum to the total population score for the facility.

The tight connection between the high-resolution toxicity-weighted concentration estimates of potential chronic human health hazard from industrial pollution and small-area socioeconomic data on residents allows the production of environmental justice (EJ) ratios that expresses the total RSEI exposure for people within a demographic category of interest, for example people with income under the US poverty line, divided by RSEI exposure for the entire population from the same source. Numbers of people affected times exposure concentration times toxicity weight can be aggregated over multiple pollution releases, so these EJ ratios can be created for any summative entity: states, cities, parent companies, and so on (Ash and Boyce 2011).

The RSEI-GM data include four different routes of public exposure to air pollution from industrial facilities: (1) direct releases to the air from point sources at the facility, such as smokestacks; (2) "fugitive" releases from undetermined points at the facility, such as open storage containers or spills; (3) releases of chemicals not destroyed by incineration that occur after transfer of the chemical from the originating facility to an incineration facility; and (4) transfers of chemicals by public or private sewerage to publicly operated treatment works (POTWs) resulting in air emissions from the volatilization of the chemical from the POTW. RSEI tracks the transfer of chemicals from TRI facilities to treatment facilities, that is, incinerators and POTWs, and models the transfer sites as the source of release within the RSEI-GM grid. For purposes of estimating receptor concentrations, this tracking follows chemicals into the landscape by introducing source locations that are not necessarily themselves TRI facilities. Responsibility for these offsite releases is assigned to the TRI facility that originally produced and transferred the chemical.

There are consistency issues in data analysis of RSEI-GM data that researchers should take into account: late revisions to TRI data, regulatory changes in TRI reporting rules, and variation in chemical speciation for TRI chemicals (notably in the case of chromium). Late revisions to TRI data occur because EPA allows TRI reporters to revise past data submissions at any time: these can be corrected in RSEI by multiplying scores by the ratio of the new to old release amount (only downward, since upward corrections often would involve creating a new score where none existed). Variation in the methods used to calculate RSEI scores can occur from both changes in the RSEI estimation methods and changes in the regulations that require TRI reporting. To compare RSEI scores across TRI data years, these changes have to be removed by using a "core chemical data set" or "core industry set" that excludes chemicals or industries whose reporting requirements have changed across the relevant range of years. Lastly, some of the chemicals that are reported to TRI are actually groups of chemicals rather than single chemical entities. For instance, chromium is reported to TRI as either chromium or chromium compounds and is modeled by RSEI as a single chemical category. However, hexavalent chromium Cr(VI) and trivalent chromium Cr(III) have very different human health risks, and the researcher may need to look into RSEI's chromium speciation estimates in detail.

Parent assignment: corporate research to assign TRI facilities to ultimate owners

The process described earlier can be used to compute EJ ratios by geographic area, facility, industry, chemical, and location, but not by parent company. The latter information is important, however, since corporate policy can affect the severity of environmental justice disparities, and finding out a corporation's total and comparative responsibility is one of the tools that communities sometimes use when they try to make political change. An early social-scientific analysis of TRI (Hamilton 1995) demonstrated that financial markets respond with reduced valuations to information about companies with facilities represented in the TRI, for example, because shareholder estimates of legal liability may be higher when EPA publishes new toxics information.

The logic of EPA's facility-level data collection effort is that monitoring, regulation, and enforcement are facility-level responsibilities. However, beginning with the Hamilton (1995) analysis, public and private decision makers have observed the value of relating environmental performance to corporate policy and corporate responsibility. By joining data on facility activity with data on corporate ownership, it becomes possible for socially responsible investors, corporate environmental managers, regulators, and activists to associate corporate policy and environmental activity. Both facility-level regulations, for example the requirement of filter or scrubbers in polluting industrial processes, and systematic regulation of the owner can contribute to improving the environment.

Although EPA collects parent company information and has made some effort to standardize names in reporting of parent companies, its most generally used data distribution method, the TRI National Analysis, does not feature parent company analysis. For instance, the 2017 TRI National Analysis displays data by release and transfer type, geographic location, chemical, and industry, but the only apparent place where it breaks out the data by parent company is in the source reduction and pollution prevention section, which describes generally beneficial activities. Similarly, the EPA Envirofacts TRI Basic Search allows search by facility name, geography, industry, or chemical but not corporate parent. These are data that exist within the TRI database, but they are not generally presented by EPA in the context of responsibility for pollution.

The TRI database, in principle, contains a field for parent company information, but the ownership data are not generally reliable. There are three kinds of problems: (1) the parent company may be left blank, reported inaccurately, reported with variant spelling (as there is no standardized company identification code), or reported as a subsidiary owner rather than the ultimate parent; (2) a facility may be jointly owned by more than one parent company; and (3) a facility that has changed hands may fail to update the parent company.

We regularize parent names to reduce variation. For exchange-traded companies, the nonprofit CorpWatch provides access to a US Securities and Exchange Commission database linking subsidiaries to parents. These automated database methods improve facility-parent matching, but gaps remain. For many facilities, we use Web or library searches or contact the technical contact listed on the Form R to ascertain the parent company. These time-intensive procedures improve the quality of matches. For facilities that are jointly owned by multiple parents, we assign the pollution from the facility to the majority owner. In the case of 50/50 joint ventures, we divide the pollution from the facility between the parent companies.

Many facilities and companies change hands over time. We research mergers, acquisitions, and divestments to update facility ownership data. Ownership of individual facilities can be affected by sales of specific assets or entire companies. We establish a contemporary snapshot of ownership and assign current and historical pollution to the current owner on the principle that ownership includes responsibility for the past pollution.

Environmental justice ratios: measuring corporate environmental justice performance

The Table 22.1 shows results from the Toxic 100 Air project for 2017 (the latest data year available at the time this was written) for five parent companies that rank high for disproportionate

Table 22.1 Five selected records from PERI's Toxic 100 Air Polluters Index, 2017

Company	*EJ: minority share*	*EJ: poor share*	*Toxic 100 Air rank*	*RSEI score*	*Share of score from top facility*
Chevron	76%	20%	40	807,162	68%
Schlumberger	75%	25%	96	189,054	88%
Goodyear Tire & Rubber	74%	22%	62	368,213	65%
TMS International	74%	29%	12	3,285,626	53%
Ecolab	71%	16%	79	265,619	57%

chronic human health risk to minority groups. The Toxic 100 companies are chosen from a list of major companies (as defined by them being on various Forbes, Fortune, or S&P 500 lists); this table has been further limited to those companies with less than 90% of their risk score from a single facility. The Toxic 100 rankings are on the Web, along with the underlying data for all companies in the TRI database. In the following table, "EJ: minority share" is the share of the total population health risk borne by minority racial/ethnic groups, and "EJ: poor share" is the share borne by people living below the poverty line. For comparison, in the US population, approximately 39% were members of minority racial/ethnic groups and approximately 13% lived below the poverty line in 2017.

Examination of the individual facilities for the companies listed in Table 22.1 shows especially high burdens on minority communities, the EJ: minority share, at sites in El Segundo, California; Richmond, California; Houston, Texas; Beaumont, Texas; Gary, Indiana; East Chicago, Indiana; and Fresno, Texas. These data, which integrate pollution releases, the social distribution of pollution releases, and the ultimate corporate responsibility for the exposure can intervene in public and private decision-making in several ways. First, the data connect environmental justice to corporate decisions and show how corporate policy is expressed through siting decisions and the management of facilities. Second, the publication of these data showing both the facilities and parents can connect multiple communities affected by separate facilities with common ownership, with the potential to identify patterns in company relationships with disadvantaged communities. These connections may also be useful to regulators and to socially responsible investors who can use the tool to coordinate environmental and social corporate governance (ESG).

Linking local and global pollutants: the Greenhouse Gas Reporting Program

In 2008, Congress directed the EPA to use its existing authority under the Clean Air Act to develop a mandatory greenhouse gas reporting rule, intended to cover both upstream production of fossil fuels from suppliers and downstream sources of GHGs that were large, fixed facilities (excluding mobile sources, agriculture, residential, etc.) The supplier information is useful but not immediately applicable to EJ studies since it contains locations of production rather than release. The downstream information consists of annual reports of greenhouse gas emissions from facilities in certain industries, primarily large facilities releasing 25,000 metric tons or more of CO_2-equivalent emissions (including CO_2, methane, nitrous oxide, and some fluorinated gases). The first reports were for data collected in 2010: a number of additional industries were added in 2011. Downstream reports include nearly all emissions from electricity generation and most emissions from industrial facilities, accounting altogether for about half of all US GHG emissions.

Stated justifications for the creation of the GHGRP database generally do not include explicit right-to-know language but do include general informational purposes. For instance, EPA's FAQ page (updated September 23, 2019) on GHGRP describes the benefits of the data as follows:

> Information in the database can be used by communities to identify nearby sources of greenhouse gas emissions, help businesses track emissions and identify cost- and fuel-saving opportunities, inform policy at the state and local levels, and provide important information to the finance and investment communities.

While greenhouse gases have global effects on anthropogenic climate change, the research discussed here has to do with local human health effects from breathing co-pollutants from combustion of fossil fuels such as particulate matter, NO_X, and volatile organic compounds. Co-pollutant emissions are not directly reported in the GHGRP database, but GHG emissions can be used either as a proxy or as a link to direct estimates of these emissions from another EPA source such as eGRID (Emissions & Generation Resource Integrated Database), although those databases usually cover only the electric power generation industry.

Extending EJ to greenhouse gas emitters

GHGRP source emissions have associated lat-long points and therefore can be related to US Census American Community Survey five-year data. In calculating EJ ratios from the GHGRP database, we made certain simplifying assumptions.

First, we assumed that co-pollutant severity was proportional to CO_2-equivalent emissions from fossil fuel combustion. This assumption could be improved upon in future work by treating different fossil fuels as having different co-pollutant profiles. We also omitted biogenic CO_2-equivalent emissions from the total because they are excluded from most global climate change reporting on the basis that they are not a net source of CO_2 in the atmosphere over the medium term, yet this does not prevent them from producing local co-pollutants. Second, we assumed that demographic composition of populations affected by co-pollutants could be modeled as those living within a 10-mile radius of the facility releasing them, since there is no equivalent of RSEI for the GHGRP database that does detailed exposure modeling at the facility level.[2] Total populations affected by each facility were taken as those within census blocks whose centroids were within 10 miles of the facility point location. For parent companies, the 10-mile radius population around each facility was weighted by the facility's CO_2-equivalent emissions, and these were aggregated for all facilities it owns.

Some CO_2-equivalent emissions are "non-direct emissions": for example, oil and natural gas producers report their emissions from operations within geologic basins, and distribution companies report emissions that take place over their distribution system within a state. Because these emissions do not come from point sources, they are excluded from this analysis.

Assigning responsibility for greenhouse gas emissions

As with TRI, the GHGRP database contains parent company information, but this information is not displayed by EPA as a summed-up table in its Data Highlights default public data presentation. EPA's Envirofacts and FLIGHT database do allow searches by parent company name. As with TRI, there is no overall parent company ID. Unlike TRI, the GHGRP database allows reporting of multiple parent company owners for individual facilities, instead of a single parent company, and includes percentages of ownership for each.

In general, this permits a parent company assignment procedure similar to that described earlier for TRI: attempting automatic regularization of parent names and SEC filing lookup, final decision informed through Web searches done by a researcher, facility ownership either assigned to the majority owner or to two 50%/50% owners, facility ownership assigned to most recent owner, and so on.

However, determining parent company ownership presents a few challenges that are particular to the electric power generation industry, which is the largest single sector for emissions in the GHGRP database. Ownership determination is also sometimes complicated because the facility may be named in connection with its operating company, rather than with its owner. In addition, in this industry sometimes different owners own different power generating units at the same overall facility, resulting in cases in which a facility has no 50% owner. For this reason, a number of single facilities have large enough CO_2-equivalent emissions to make the top 100 list of "companies," but are treated, in effect, as a parent company that consists of that single facility. These single-facility emissions could be divided up and assigned to other companies by percentage of ownership, but this would be somewhat problematic since percentage of ownership may not equate to percentage of the facility's emissions generated.

GHGRP EJ ratio analysis for parent companies

Table 22.2 shows results from the "Greenhouse 100" project for 2017 (the latest data year available at the time this was written) for the five parent companies that rank highest for disproportionate modeled co-pollutant exposure to minority groups. The Greenhouse 100 project and its underlying data are publicly available on the Web for all companies in the GHGRP database. Again, "EJ: minority share" refers to the share of the total population health risk borne by minority racial/ethnic groups, and "EJ: poor share" is the share borne by people living below the poverty line.

Lessons on linkages

Some EJ analyses are best done not for parent companies but for individual facilities. For these purposes, it is often helpful to link facility data from multiple sources together. In connection with our research, projects have been undertaken linking TRI to GHGRP facilities; GHGRP

Table 22.2 Five selected records from PERI's Greenhouse 100 Index, 2017

Company	*EJ: minority share*	*EJ: poor share*	*Greenhouse 100 rank*	*2017 CO_2-equivalent emissions (metric tons)*	*Share of emissions from top facility*
San Antonio Public Service Board	78%	19%	48	12,839,604	47%
LyondellBasell	77%	19%	91	7,602,442	26%
BP	74%	20%	39	15,185,278	31%
Hilcorp Energy	73%	22%	95	6,928,281	32%
Enterprise Products Partners	73%	21%	80	8,518,255	17%

Examination of individual facilities owned by these companies reveals EJ: minority share scores to be dominated by facilities in San Antonio, Channelview, Houston, Corpus Christi, and Mont Belvieu (all in Texas), and a few other locations including Whiting, Indiana, and Bloomfield, New Mexico.

to the US EPA's eGRID (for comparison of emissions with power generation, co-pollutants, and fuel quantities); and TRI to US Equal Employment Opportunity Commission (EEOC) data to compare employment of members of minority groups at facilities with the share of environmental burdens borne by members of the same minority groups living near these facilities.

These comparisons require addressing a number of difficulties with the design and accuracy of data. In some cases, the unit of data collection may focus on different managerial or engineering concepts. For example, the unit of observations for the EEOC data on employment by race, sex, and occupation considers the "establishment," an economic concept, while the TRI collects data on "facilities"; these entities often coincide, but not always. Other reporting systems, especially those concerned with energy production and industrial processes, can be based on specific activities or processes, with for example each boiler within an electricity generating facility reporting separately.

Addresses of facilities may be recorded differently in different databases, with some facilities, for example, having no set physical address other than a point some miles down a rural road. Mailing addresses may be listed instead of physical addresses, which if uncorrected could lead to pollution being attributed to a corporate headquarters rather than a physical plant. Some of these problems can be mitigated by using GIS or other methods of comparing lat-long coordinates, although these are sometimes missing, incorrect, or poorly defined (as when a facility with a large physical extent must be reduced to a single point). As with parent company assignments, we have found no better way to make these matches than to have them automatically suggested by programs as far as possible, with a researcher making the final decision.

The US EPA constructed an additional database, the EPA Facility Registry Service (FRS) that is intended to assist with linkages across EPA datasets to facilitate comparisons and correlations. FRS succeeded previous EPA internal systems that were intended to solve a fundamental informational and regulatory problem: EPA has different programs authorized under different laws with a host of differing definitions. Although all of the regulations refer to facilities in some sense, definitions of what constitutes a facility may differ. Even in cases where the definitions largely correspond, data are collected by each of these programs independently, without any mandate for any agent or regulator to figure out whether, for example, a facility with an air pollution permit under the Clean Air Act is the same facility as one listed with a hazardous waste permit in the Resource Conservation and Recovery Act (RCRA) Information System.

FRS assigns a single ID to each facility that EPA has identified, and attaches this ID to all of the separate air, water, hazardous waste, and so on IDs that the same facility has under various EPA programs. There may be zero, one, or many IDs for each EPA program that are associated with a single FRS facility ID. The FRS database can be obtained through EPA's Envirofacts website, and we recommend it as the starting point for any kind of database-to-database facility comparison using EPA data.

In the case of the join of TRI/RSEI data with EEOC data (Ash and Boyce 2018), the join that permitted the analysis of jobs and pollution required matching two completely independent sets of identifiers, that of TRI/RSEI and that of EEOC. The set of facilities targeted for the join was limited to the 1,000 highest-impact facilities in terms of RSEI score, out of approximately 20,000 reporting facilities. The join, based on matching name and address, succeeded in joining more than 700 facilities. Walker (2013, online appendix A.4) describes matching rates using name and address to join data between the Census Bureau Standard Statistical Establishment List (SSEL) and US EPA facility lists. A key distinction is that, unlike the TRI right-to-know data that are publicly available and specifically and intentionally identify facilities, the EEOC and Census Bureau datasets, while collected by government agencies with a public mandate, are confidential, and access and use are tightly restricted. Access is limited to research by stringent application, and only summary results and generalized findings may be reported. For example, the access to EEOC data for Ash and Boyce (2018) required formal appointment of the investigator as an (unpaid) employee of the EEOC, demonstration that the research would contribute specifically to EEOC meeting its agency mandate, and strict regulation that no individual facility data be released.

Other datasets on firm and facility activity are proprietary with expensive access, for example, the Compustat dataset on the financial and real activity of firms traded on stock exchanges. In many cases, the right to publish data about specific firms is limited by user agreements for proprietary databases. The interface of right-to-know data with other datasets is of potentially great value, but the usefulness of right-to-know data is curtailed when joined datasets are proprietary or otherwise restricted in access and results are limited to aggregated summaries and general findings.

The web of connections to other databases can also expand value. The Toxic 100 and Greenhouse 100 indexes link to public and non-governmental databases on chemical toxicity, to additional facility-level information maintained by the US EPA, to mapping services provided by private providers, and to several public watchdog databases maintained by the non-governmental organization Good Jobs First.

Applied and social scientific research

Once the data have been assembled from the multiple sources described previously, they can be used in a variety of overlapping research projects. This section describes some of the uses that have been made of these data generated by researchers centered at the University of Massachusetts Amherst.

Bouwes et al. (2003) and Ash and Fetter (2004) pioneered the application of RSEI to environmental justice. In both studies, the unit of observation is the geographic receptor – in the case of Bouwes et al., the RSEI square kilometer cell and in the case of Ash and Fetter, the census block group. The dependent variable is human health risk, and the key explanatory variable is the minority share of the population. Important methodological differences between Bouwes et al. (2003) and Ash and Fetter (2004) include the assessment of all areas in Bouwes et al., as opposed to urbanized areas in Ash and Fetter, and the inclusion of population-weighted risk score in Bouwes et al., as opposed to the analysis of individual unit risk in Ash and Fetter.

Both studies found substantial evidence of environmental inequality on racial and ethnic lines. An enormous advantage to the high-resolution modeling of fate and transport of pollution is that it obviates the need to debate "how close is close" that plagued earlier studies based simply on proximity to a polluting facility (see Mohai and Saha 2006, 2007 for discussion of these problems). The comprehensive receptor-based modeling of TRI data with RSEI enabled analysis that was both national in scope and precise regarding exposure.

The high geographic resolution of the RSEI model enables the analysis of neighborhood-level differences in exposure to industrial toxics. Ash and Fetter focused on within-city risk differences, comparing this to overall (pooled within- and between-city) differences in risk. Given the importance of residential segregation in US cities, local siting decisions by companies, and local regulatory enforcement, the focus on distribution of industrial toxic exposure within urban areas allows Ash and Fetter to pose the question, "Who lives on the wrong side of the environmental tracks?"

The distinction between within-city and between-city differences in exposure provided new information on the disproportionate exposure of Latinos, or Hispanics, to industrial toxics in the United States. Earlier research had focused on and identified disproportionate exposure of

African Americans to industrial toxics, with disproportionate exposure occurring on essentially every geographic scale, both neighborhoods within cities and excess exposure based on population concentration in US industrial cities in the industrial Midwest and the urban South. Latinos were more concentrated in parts of the US with less toxic-intensive heavy industry, and city-level comparisons did not identify disproportionate Latino exposure. However, neighborhood-level analysis within cities demonstrated that Latinos live in parts of cities that have systematic excess exposure (Ash and Fetter 2004). Case studies of specific regions, for example the analysis by Morello-Frosch et al. (2001) of the "riskscape" of the Los Angeles basin, had detected this phenomenon, and RSEI-based national analysis confirmed its systemic character of disparities within place.

The high-resolution RSEI-GM data can support hierarchical models that examine simultaneously the distribution of pollution within a polity, which requires high-resolution distinction among neighborhoods, and the overall level of average pollution in that polity compared to others. Building on Ash and Boyce (2011) and Ash et al. (2009), which developed an empirical measure of the segregation of pollution, Ash et al. (2013) tested a political economy model in which the degree of environmental disparity, that is, the capacity to displace pollution onto a vulnerable social group, affects the overall level of pollution in metropolitan areas. This operationalizes Boyce's work on the theory of inequality and environmental degradation, which hypothesizes that the ability to displace environmental bads onto vulnerable populations (into spaces that effectively become "sacrifice zones") and to appropriate environmental goods into spaces reserved for a privileged few affect the political calculus regarding environmental bads and goods. Ash et al. (2013) find that in high-disparity metropolitan areas, not only do vulnerable social groups, including people of color and low-income people, experience substantially higher pollution exposure, but also the overall level of pollution exposure is higher.

The high geographic resolution of the Geographic Microdata also makes it possible to compare environmental justice gradients, the extent to which vulnerable communities are disproportionately exposed, across locations. For example, the states of the industrial Midwest – Illinois, Indiana, Michigan, Minnesota, Ohio, and Wisconsin (together designated as EPA Region 5), have high exposure of the average resident and also a very steep gradient, in which racial and ethnic minorities are disproportionately exposed (Zwickl et al. 2014). It is also possible to compute vertical inequality measures that describe the variation in exposure between the most exposed and least exposed communities, and to compare these to horizontal inequality measures based on differential exposure by race or class (Boyce et al. 2016). Currie et al. (2015) used variation in RSEI scores to value environmental health risks from changes in housing values induced by plant openings and closings.

More recent research involves the integration of Risk Screening Environmental Indicators data with carbon emissions data from the Greenhouse Gas Reporting Program. Boyce and Pastor (2013) drew attention to the importance of explicitly considering air quality co-benefits and environmental justice in the design of carbon policy. Subsequent work, including Cushing et al. (2018) and Boyce and Ash (2018), have expanded the global-local analysis of greenhouse gas reductions, the potential for co-benefits, and the peril of overlooking co-benefits for environmental justice communities. The combination of RSEI and GHGRP data makes it possible to explore further the ways in which this can be achieved.

Public intermediation for policy impact

In addition to research uses, these data have been used in various information intermediation efforts, with purposes including regulatory compliance and enforcement, socially responsible investment, corporate environmental management, and popular and mass movement awareness, action, and redress. Typically, these projects have taken the form of a website allowing the public to search and display the data for free, in an attempt to empower one or more of these types of uses. This kind of activity has taken place in connection with TRI since 1989, before the World Wide Web was created, on Bulletin Board Systems and through other early means of networked data sharing and display.

The US EPA itself has created search-and-display sites, which generally also include data download and mapping capabilities, for disseminating the TRI and GHGRP data. The major EPA sites at the time of writing are TRI Explorer, EnviroFacts (which contains both TRI and GHGRP as well as many other EPA databases), and FLIGHT (a GHGRP interface). Other governmental sites that distribute these data include international sites focused on PRTRs (pollutant release and transfer registries) that include TRI along with similar data from other countries. This is done by OECD (the Organisation for Economic Co-operation and Development) and UNITAR (United Nations Institute for Training and Research), and has been done by the CEC (the North American Commission for Environmental Cooperation, established under the North American Agreement on Environmental Cooperation).

Nonprofit, journalistic, and academic organizations have operated websites to increase public access to the data and enable analyses that are difficult to undertake on the official sites. One of the earliest efforts at enhancing public access to TRI and other EPA databases was RTK NET (the Right-to-Know Network), a project of the nonprofit Center for Effective Government (previously named OMB Watch). RTK NET has provided access to TRI and other databases since 1989. After the Center for Effective Government closed in 2016, the *Houston Chronicle* newspaper sponsored RTK NET. Another notable site was Scorecard, which provided a value-added interface to TRI and related exposure data prior to the advent of RSEI. Scorecard was initiated by Environmental Defense (formerly the Environmental Defense Fund), a major environmental advocacy nonprofit. GoodGuide.org temporarily operated as a legacy site that maintained but did not update the Scorecard data.

The Toxic 100 and Greenhouse 100 are public data intermediation projects run by the Corporate Toxics Information Project of the Political Economy Research Institute (PERI) at the University of Massachusetts Amherst, a public university, as part of its public mission to engender greater public participation in decision-making about environmental policy. These lists rank US companies by their emissions responsible for global climate change, by chronic human health risk from air toxics exposure, and by chronic human health hazard from water pollution exposure. The PERI analysis also includes environmental justice indicators, examples of which were given earlier, to assess impacts on minorities and low-income people. These indexes are frequently cited in news media, on Wiki pages about individual corporations, and in shareholder resolutions on corporate environmental performance.

Several additional intermediation projects have in turn used the Toxic 100 index to add further value to the pollution information from the Corporate Toxics Information Project. In 2008, the UK-based Business and Human Rights Resource Centre moderated a dialogue between the top 10 firms listed on the Toxic 100 index published that year and the Corporate Toxics Information Project of PERI (Business and Human Rights Resource Centre 2008). Good Jobs First, a non-governmental policy resource center and the Corporate Toxics Information Project reciprocally link company specific data between the Toxic 100 and Greenhouse 100 and the Violation Tracker and Subsidy Tracker sites, which monitor and report fines and penalties that corporations pay for violation of environmental, health, occupational, financial, and fiscal regulations and laws and federal, state, and local public subsidies to corporations.

Conclusion

To affect environmental justice policy and practice, the right-to-know movement and regulation by right to know require consistent intermediation by public, university, and non-governmental organizations to draw out meaningful connections in the data. The Corporate Toxics Information Project has experimented with data interlinkages and the intermediation of results to expand the impact and utility of right-to-know data from pollutant release and transfer registers. Concrete scientific findings include results on the political economy of pollution exposure and environmental racism in the United States and the weak empirical case for a widely assumed jobs-environment trade-off. Public intermediation of the data has affected shareholder intervention in the dimension of socially responsible investment and activist and journalistic interventions. Effective interlinkage and intermediation depend on the availability of data for integrative assessment models, the establishment of corporate ownership of fixed assets, and socioeconomic variation on a geographic basis. Establishment of common identifiers at the facility, corporate, and geographic level is a significant challenge for environmental justice researchers.

Notes

1 Who's Counting? The Systematic Underreporting of Toxic Air Emissions. Environmental Integrity Project. June 2004. www.environmentalintegrity.org/pdf/publications/TRIFINALJune_22.pdf

2 The selection of the 10-mile radius reflects an expert judgment on the most affected area for many TRI releases in the RSEI fate-and-transport model. Criticisms of distance-based buffer models include Mohai and Saha (2003), Mohai and Saha (2006, 2007).

References

Ash, M. and Boyce, J.K., 2011. Measuring corporate environmental justice performance. *Corporate Social Responsibility and Environmental Management*, 18(2), pp. 61–79.

Ash, M. and Boyce, J.K., 2018. Racial disparities in pollution exposure and employment at US industrial facilities. *Proceedings of the National Academy of Sciences*, 115(42), pp. 10636–10641.

Ash, M., Boyce, J.K., Chang, G., Pastor, M., Scoggins, J. and Tran, J., 2009. *Justice in the air: Tracking toxic pollution from America's industries and companies to our states, cities, and neighborhoods*. Amherst, MA: Political Economy Research Institute.

Ash, M., Boyce, J.K., Chang, G. and Scharber, H., 2013. Is Environmental justice good for White folks? Industrial air toxics exposure in Urban America. *Social Science Quarterly*, 94(3), pp. 616–636.

Ash, M. and Fetter, T.R., 2004. Who lives on the wrong side of the environmental tracks? Evidence from the EPA's risk-screening environmental indicators model. *Social Science Quarterly*, 85(2), pp. 441–462.

Bouwes, N.W., Hassur, S.M. and Shapiro, M.D., 2003. Information for empowerment: The EPA's risk-screening environmental indicators project. In *Natural Assets: Democratizing Ownership of Nature*. Eds. J.K. Boyce and B. Shelley. Washington, DC: Island Press, pp. 117–134.

Boyce, J.K. and Ash, M., 2018. Carbon pricing, co-pollutants, and climate policy: Evidence from California. *PLoS Medicine,* 15(7), e1002610.

Boyce, J.K. and Pastor, M., 2013. Clearing the air: Incorporating air quality and environmental justice into climate policy. *Climatic Change*, 120(4), pp. 801–814.

Boyce, J.K., Zwickl, K. and Ash, M., 2016. Measuring Environmental Inequality. *Ecological Economics*, 124, pp. 114–123.

Business and Human Rights Resource Centre, 2008. Toxic 100 & company responses – 2008. www.business-humanrights.org/en/documents/toxic-100-company-responses-2008.

Currie, J., Davis, L., Greenstone, M. and Walker, R., 2015. Environmental health risks and housing values: Evidence from 1,600 toxic plant openings and closings. *American Economic Review*, 105(2), pp. 678–709.

Cushing, L., Blaustein-Rejto, D., Wander, M., Pastor, M., Sadd, J., Zhu, A. and Morello-Frosch, R., 2018. Carbon trading, co-pollutants, and environmental equity: Evidence from California's cap-and-trade program (2011–2015). *PLoS Medicine*, 15(7), p.e1002604.

Fung, A., Graham, M. and Weil, D. 2007. *Full disclosure: The perils and promise of transparency*. New York: Cambridge University Press.

Hamilton, J.T., 1995. Pollution as news: Media and stock market reactions to the toxics release inventory data. *Journal of Environmental Economics and Management*, 28(1), pp. 98–113.

Hersh, A. 2006. Corporate Toxics Information Project Interview with co-directors James Boyce and Michael Ash. Political Economy Research Institute. March. www.peri.umass.edu/images/CTIP_Interview.pdf.

Mohai, P. and Saha, R., 2003, August. Reassessing race and class disparities in environmental justice research using distance-based methods. In *Annual Meeting of the American Sociological Association*, Atlanta, GA.

Mohai, P. and Saha, R., 2006. Reassessing racial and socioeconomic disparities in environmental justice research. *Demography*, 43(2), pp. 383–399.

Mohai, P. and Saha, R., 2007. Racial inequality in the distribution of hazardous waste: A national-level reassessment. *Social Problems*, 54(3), pp. 343–370.

Morello-Frosch, R., Pastor, M. and Sadd, J., 2001. Environmental justice and Southern California's "riskscape" the distribution of air toxics exposures and health risks among diverse communities. *Urban Affairs Review*, 36(4), pp. 551–578.

Walker, W.R., 2013. The transitional costs of sectoral reallocation: Evidence from the clean air act and the workforce. *Quarterly Journal of Economics*, 128(4), pp. 1787–1835.

Zwickl, K., Ash, M. and Boyce, J.K., 2014. Regional variation in environmental inequality: Industrial air toxics exposure in US cities. *Ecological Economics*, 107, pp. 494–509.