Article

AMEND: Open Source and Data-Driven Oversight of Water Quality in New England

Nathan Edward Sanders

Mystic River Watershed Association, Arlington, MA 02476, USA; E-Mail: openamend@gmail.com

Submitted: 29 March 2019 | Accepted: 15 May 2019 | Published: 6 August 2019

Abstract

The advent of government transparency through online data publication should provide a transformative benefit to the information gathering practices of civic organizations and environmental advocates. However, environmental agencies and other reporters often disseminate this critical data only in siloed repositories and in technically complex, inconsistent formats, limiting its impact. We have developed a new open source web resource, the Archive of Massachusetts Environment Data or AMEND, which curates information relating to federal, state, and local environmental stewardship in Massachusetts, focused on water quality. We describe the construction of AMEND, its operation, and the datasets we have integrated to date. This tool supports the development and advocacy of policy positions with published analyses that are fully reproducible, versioned, and archived online. As a case study, we present the first publicly reported analysis of the distributional impact of combined sewer overflows on Environmental Justice (EJ) communities. Our analysis of the historical geospatial distribution of these sewer overflows and block-level US Census data on EJ indicators tracking race, income, and linguistic isolation demonstrates that vulnerable communities in Massachusetts are significantly overburdened by this form of pollution. We discuss applications of this analysis to the state-level legislative process in Massachusetts. We believe that this approach to increasing the accessibility of regulatory data, and the code underlying AMEND, can serve as a model for other civic organizations seeking to leverage data to build trust with and advocate to policymakers and the public.

Keywords
advocacy; databases; environmental data; environmental justice; Massachusetts; open source; policy; water quality

Issue

This article is part of the issue “Civic Organizations and Digital Technologies in an Age of Distrust”, edited by Eric Gordon (Emerson College, USA).

© 2019 by the author; licensee Cogitatio (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

1. Introduction

Establishing trust between policymakers, civic organizations, and the public about the merits of policy decisions requires agreement on the facts underlying policy issues. That agreement is predicated not only on the existence of robust data on present and historical conditions and government actions, but also on shared access to that data. As Janssen, Charalabidis and Zuiderwijk (2012) wrote of the benefits of government data transparency, “By opening data, users can validate and verify whether the conclusions drawn from the data are correct and justified, and they can analyze the previously collected data to sharpen the focus of policy-making.” The advent of government transparency through online data publication should provide a transformative benefit to “interaction between governments, citizens, and the business sector” and especially the information gathering practices of civic organizations and environmental advocates (Bertot, Gorham, Jaeger, Sarin, & Choi, 2014). Particularly as governments move towards more data-driven administrative procedures, data transparency will become an increasingly important aspect of democratic accountability (Redden, 2018).

Access to data at scale on American regulatory policy and enforcement has enabled a variety of impactful academic research in recent years on topics including the relationship between regulation and innovation...
(Jaffe & Palmer, 1996), the inter-connectivity between states’ environmental regulatory policy (Konisky, 2007), how border-adjacency incentivizes free ridership in regulating air pollution (Konisky & Woods, 2010), and much more. Meanwhile, a variety of web portals and online tools such as the US Geological Survey and US Environmental Protection Agency’s (EPA) Water Quality Portal (National Water Quality Monitoring Council, n.d.) are used to publish and distribute particular datasets useful for this field of research. Together, this set of tools supports experienced researchers with technical skillsets in undertaking substantial research studies. But, in general, the adoption of open data policies among governments worldwide has been slow and uneven and the provision of tools to make public data truly accessible has lagged behind the mere publication of data (Bertot et al., 2014; Janssen et al., 2012). The studies cited above have generally required significant querying, reformatting, transformation, and cleaning of data, and often have required the integration of data across several sources.

Moreover, we have experienced that the existing set of tools is not supportive of stakeholders with less time, money, or technical expertise. These users, including policymakers, civic organizations, journalists, and citizen advocates, have difficulty identifying, accessing, and using data resources. Meanwhile, the diversity and scale of relevant, publicly available data is expanding due to new technologies and reporting requirements. All these factors hamper stakeholders’ ability to do analysis that would inform their actions on the policy landscape. These challenges are roadblocks to incorporating public data into the policy oversight and advocacy role of these stakeholders; they limit shared access to a common set of facts.

We believe that an ideal data resource for civic organizations should meet the following requirements to promote reproducibility, extensibility, and trust by users:

- Serves diverse users: The resource should increase the accessibility of relevant data to stakeholders from communities including civic organizations, journalists, and citizen advocates, including users that have and have not previously worked with the constituent datasets and users who do and do not have technical backgrounds.
- Transparent: The origin of, and history of changes to, all data and analysis elements provided on the resource should be clearly stated, versioned, and traceable.
- Open source: The resource should be reliant only on open source technologies to develop and maintain the site. The barrier for other contributors to participate in the development process should be low and collaboration should be encouraged.
- Low development and maintenance cost: It must be possible to create, host, and keep up the resource without dedicated web development personnel, e.g., with a single volunteer or small part-time team.

These principles align to the tenants of the movement for “reproducible science” (Peng, 2011; Schwab, Karrenbach, & Claerbout, 2000). When research is reproducible, “all details of the computations—the underlying data and the code that generated the results—are made conveniently available to others” (Stodden & Miguez, 2014). A culture of reproducibility not only supports other researchers in leveraging and extending past research, but also generates confidence and trust in analysis and results.

In this article, we present our work towards establishing the Archive of Massachusetts ENvironmental Data (AMEND). AMEND is an open access, integrated repository of environmental regulatory data and analysis focused on enhancing the use of evidence and accountability for water policy in Massachusetts and the New England region, which has been designed to adhere to the four principles outlined above. We describe the local context for issues related to water policy and enforcement in the region in Section 2 and detail the development and features of AMEND in Section 3. Section 4 provides a case study of the application of AMEND to a high-profile contemporary water policy issue and Section 5 contextualizes this work in terms of modern concepts in media theory. We conclude in Section 6 with an outline of future work including planned features and extensibility to other localities.

2. Local Context

Massachusetts has a long history of environmental advocacy and large-scale environmental policymaking, and a vibrant present-day community of watershed associations and other environmental groups. Massachusetts was home to the well-known lawsuit against W. R. Grace and Beatrice Foods over groundwater contamination in the city of Woburn and the ensuing Superfund cleanup (Brown, 1987; Kiel & Zabel, 2001); embarked on a trailblazing, multi-billion dollar combined sewer overflow (CSO) elimination and mitigation project to clean up the Boston Harbor and connected waterways (Dolin & Levy, 1990; Levy & Connor, 1992); and adopted climate change oriented emissions reductions a decade ago under the Regional Greenhouse Gas Initiative (Byrne, Hughes, Rickerson, & Kurgelashvili, 2007). In each of these cases, data—including data on water quality, epidemiology, effluent, and emissions—has been instrumental to motivating action as well as to monitoring and verifying the efficacy of that action.

2.1. Stakeholders

Massachusetts is home to a rich array of civic organizations dedicated to environmental protection and water resources, specifically. Together, these groups continue a centuries-long legacy of political organizing around rivers in the US (see e.g., Randolph, 2018).

The Mystic River Watershed Association (MyRWA) is a non-profit organization dedicated to the preservation
and enhancement of the Mystic River Watershed in Eastern Massachusetts. MyRWA is a science-based environmental advocacy organization that operates dedicated observational programs to study water quality, stormwater pollution, and fisheries health and educational programs to inform students and the broader community about these and other topics. Its mission is to protect and restore the Mystic River, its tributaries, and watershed lands for the benefit of present and future generations and to celebrate the value, importance, and beauty of these natural resources. The MyRWA Policy Committee is a group of staff and volunteers that collaborate to engage in advocacy in service of this mission. The Committee’s work includes filing comment letters in response to development proposals and permit applications and developing testimony on behalf of, or in opposition to, environmental legislation and rules at the local, state, and federal level. Much of MyRWA’s work is supported by governmental and foundation grants. All these written materials—grant proposals, reports, comment letters, and testimony—regularly contain references to data on water quality conditions, status of impairment in a water body, and permit or license conditions to bolster an argument about what type and level of resources or regulations are needed in a local area.

Most major water bodies in Massachusetts have active watershed groups associated with them, such as the Charles River Watershed Association, Connecticut River Conservancy, and Merrimack River Watershed. Economic research across more than 2,000 US watersheds has found that higher activity among such watershed groups in the US is a causal factor of improved local environmental water quality (Grant & Langpap, 2018). Many of these groups are members of the statewide Massachusetts Rivers Alliance. Many regional and national environmental groups also do substantial work on water policy in Massachusetts, such as the Conservation Law Foundation and Environmental League of Massachusetts.

Among the journalistic organizations covering environmental policy in Massachusetts are regional periodicals (e.g., Boston Magazine, The Boston Globe), radio networks (e.g., WBUR and WGBH radio), local newspapers (e.g., The Eagle-Tribune, Worcester Telegram), and national outlets like Inside Climate News.

While much of our work is motivated by the needs of the MyRWA Policy Committee, we view each of these organizations as potential users of AMEND. Individuals and organizations are often not aware of the variety and disparate sources of information related to their work that is published by public agencies. Creating an accessible, transparent, and centralized repository of data improves the ability of advocates to examine this information and invites a larger and more diverse base of contributors. We seek to develop AMEND as a platform to facilitate collaboration across these organizations and their constituents.

2.2. Regulatory Environment

As in any US state, a web of federal, state, and local agencies are responsible for environmental regulation in Massachusetts. Among these are the US EPA; the Department of Environmental Protection (DEP) within the Massachusetts Executive Office of Energy and Environmental Affairs (EEA); the Attorney General of Massachusetts; the Conservation Commissions of each municipality within the state; and our state legislature, the General Court of the Commonwealth of Massachusetts.

One of the motivating factors for our work was an April, 2016 declaration by MA Governor Charlie Baker to pursue delegation of the US Clean Water Act’s National Pollutant Discharge Elimination Systems (NPDES) program to Massachusetts, which would transition the Commonwealth from federal to state primacy for oversight of this important regulatory instrument (Office of Governor Charlie Baker and Lt. Governor Karyn Polito, 2016). In a unified response, the state’s environmental advocates opposed this delegation on the grounds that the DEP was already underfunded to pursue its current mandate and that a sustainable source of funding for the DEP to maintain staffing on oversight in future years had not been identified (Abel, 2016). Data on the historical funding and staffing levels of the DEP, and how agency outcomes like enforcement actions relate to those resources, were instrumental in providing evidence to evaluate and support argumentation around this issue. The proposal was defeated and, when the Governor reintroduced a similar bill in the next session (Baker, 2017), an expanded effort drawing on these data sources was again successful in defeating delegation.

MyRWA also evaluates and publicly comments on Massachusetts General Law Chapter 91 (Ch. 91), the MA Public Waterfront Act, permit applications and renewals. Ch. 91 codifies a public trust doctrine preserving public access to coastal and inland waterways, which include much of the Mystic River and its tributaries. In reviewing these applications, there is often a need to understand the permit conditions of similar properties, which have not in general been readily available for comparison.

As a final example of the regulatory environment in MA, consider the US EPA’s General Permits For Stormwater Discharges From Small Municipal Separate Storm Sewer Systems (MS4) in Massachusetts (US EPA, n.d.-b). The permit, which ultimately took effect in July 2018, was drafted to replace a 2003 MS4 permit that expired in 2008. Because the permit imposes stronger stormwater regulations on more than 200 municipalities, it will have profound fiscal and environmental impacts throughout the state. Understanding the state of impairment and sources of pollution to water bodies in each of these municipalities should play a fundamental role to the community and government’s approach to management and oversight of this important new policy.
2.3. Existing Resources

While Massachusetts has a large number of state-specific regulations, enforcement agencies, and authorities that each generate individual data assets that can be used to understand their work, there have historically been few resources available for accessing and manipulating data related to the issues they govern. Available MA-specific resources included certain datasets published on the EEA’s website, including an employee directory and fish mercury data; MassBudget’s Budget Browser (Massachusetts Budget and Policy Center, n.d.); and US EPA’s Watershed Assessment, Tracking & Environmental Results System (US EPA, n.d.-c). Recently (August 2017), the EEA Executive Data Portal (Massachusetts EEA, n.d.) was made available and offers a web query service and visual dashboards for data related to state permits, facilities, inspections, enforcement, and drinking water measurements. The Data Portal represents a substantial step forward in state-provided data services for the stakeholder community. However, its scope is limited to certain state agency-generated data assets.

This data has not been widely used by the groups listed in Section 2.1. for reasons of awareness, ease of access, and comprehensiveness. For example, a 2017 Boston Globe article (Abel, 2017) discussing the relationship between DEP funding levels and enforcement activity relied on Freedom of Information Act requests to the agency for aggregate reporting on enforcement levels, staffing, budget, etc. rather than making use of the online data resources related to these issues. A 2019 WBUR story about the public health threat of sewage overflows (Wasser, 2019) relied on data consolidated in 2013 by other journalists (see also Section 4.3.).

In general, there are several reasons why advocates and civic organizations are motivated to develop their own data repositories. First, much of the analysis they seek to perform (as in Section 4 and other examples cited in this article) is comparative and integrative, requiring data sets published by different agencies to be brought together. Second, analyses involving manipulation of data assets such as text processing, feature engineering, or statistical modeling benefit from direct access to data and may be complicated or slowed by mediation by a third party service. Finally, as digital publication is ephemeral, maintaining an independent repository mirroring public data assets ensures that they will continue to be available (so long as the repository maintainer persists), regardless of changes in regulation or administration at public agencies.

3. Archive of Massachusetts ENvironmental Data

To address these conditions and improve the accessibility of integrated environmental regulatory and quality data for Massachusetts and the New England region, we have developed the open source and open access AMEND (AMEND; Figure 1).

3.1. Development

AMEND has been developed to adhere to the principles of a trust-promoting public data resource outlined in Section 1. AMEND is entirely open source, built us-
ing free software and hosted using low cost tools. The full list of open source tools used to construct AMEND are specified as dependencies in the codebase hosted at the AMEND repository (Sanders, n.d.); we will provide an overview here.

Development of AMEND takes place on a public GitHub repository at which any developer can inspect all source code and data associated with the project, review the history of changes to those files, contribute modifications as a pull request, or fork their own version of the site. The GitHub Pages feature oriented around the static site generator Jekyll is used to host the user-facing site. Only one element of the AMEND infrastructure, large file storage, has a direct cost. We use Google Cloud Storage to serve the integrated database file itself and other large datafiles.

Several javascript libraries are used for data interaction and plotting on the site. Chart.js is used to generate interactive line, bar, and scatter plots and Leaflet is used to display interactive maps. The library sql.js is used to enable interactive querying of the site’s integrated database and MathJax is used to display mathematical formulas.

Multiple tools are used for data pre-processing. Tabula was used to extract tables from PDF files. The python libraries numpy and pandas are used for numeric data manipulation and analysis and pystan (Stan Development Team, 2018) is used to fit statistical models. MapShaper was used to convert town and watershed shapefiles into simplified polygons in geo-json format for efficient web display.

Upkeep of AMEND generally requires little maintenance. A single shell script is used to refresh the data sources integrated into the site and update all associated web pages and analyses. This procedure is vulnerable to changes in each of the data source repositories, for example HTML changes in scraped web pages and deprecation of API functions. As a result, some modification to data acquisition scripts may occasionally be required, and the refresh script cannot be automated to run on a schedule unless data source testing is also automated.

3.2. Features

The site is organized around three primary features:

- Data: An overview of the integrated database and individual pages describing each constituent data source with sample data tables and high level visualizations of the data.
- Analysis: Pages with analyses illustrating how to query, combine, and extract insight from the data within the integrated database. Each analysis page features descriptions of relevant findings each supported by interactive visualizations. See Section 4 for a detailed example.
- Query: A browser-based interactive query tool for executing SQL commands against the integrated database. Some sample SQL queries are provided as examples.

3.3. Datasets

To date, we have integrated a variety of state and federal data sources, as well as data from other civic organizations, into AMEND. These include:

- Data assets from the EEA Executive Data Portal, including data on enforcements, facilities, inspections, permits, and chemical measurements for drinking water.
- Line-item level DEP budget data from MassBudget.
- Individual current and historical DEP staff records from the MA Office of the Comptroller.
- Detailed text descriptions of enforcement actions posted on the DEP website, with simple text processing performed to extract penalty amounts, municipalities, and topic information.
- NPDES permit documents and basic metadata for all states in US EPA Region 1 (New England).
- Ancillary datasets to support analysis like US Social Security Administration wage inflation data and US Census American Community Survey municipal-level population data.

Over time, we plan to continually expand the list of datasets integrated into AMEND. The addition of new data sources requires technical expertise, especially when those sources present challenges in terms of poor formatting or inconsistent syndication. However, the open source code developed for the above listed data resources provides supportive examples/templates for contributors seeking to incorporate additional data sets.

4. Analysis Case Study

The “Analysis” section of the AMEND website links to posts that illustrate the usage of the integrated datasets for policy analysis. The code used to generate each analysis from the AMEND database is available in the repository. In this section, we provide a detailed overview of our analysis of the distributional impacts of sewage overflows in Massachusetts. Visitors to the AMEND website can also find analyses of the impacts of declining DEP budgets on the agency’s staff capacity and experience level and the correlation between state budgets and the volume and scope of enforcement actions undertaken by the agency.

The analysis of this section—and all others published on the AMEND website—can be reviewed, reproduced, modified, and extended by accessing the detailed statistical explanation (Sanders, 2019b) and the underlying data and code (Sanders, 2019a) published on the AMEND GitHub. These resources reduce the barrier for other stakeholders to produce their own independent analysis of the data assets integrated with AMEND to support
their own policy development and/or advocacy objectives. However, leveraging these resources and performing such analysis does require some level of technical competency and statistical knowledge.

4.1. The Environmental Justice Movement

Environmental Justice (EJ) is a global movement that seeks to create an equitable distribution of the risks, benefits, and decision-making power associated with environmental pollution, especially as these factors affect vulnerable communities (see Brulle & Pellow, 2006; Schlosberg, 2009, for recent reviews). In Massachusetts, an equal right to environmental protection is enshrined in the state constitution (Article XLIX; amended 1972), yet substantial inequalities persist across communities in the Commonwealth into the twenty-first century (Faber & Krieg, 2002).

Massachusetts has recently promulgated a new definition of EJ:

"Environmental justice is the equal protection and meaningful involvement of all people and communities with respect to the development, implementation and enforcement of energy, climate change, and environmental laws, regulations and policies and the equitable distribution of energy and environmental benefits and burdens. (Massachusetts EEA, 2017)"

This state policy identifies EJ populations according to any one of three threshold criteria applied at the Census block group level: that 65% of the households fall below the statewide median income (“Low income” criteria); that 25% or more of residents identify as non-white (“non-white” criteria); or that 25% or more of households have no member over the age of fourteen who speaks English only or very well (“English Isolation” criteria).

4.2. Combined Sewer Overflows

CSOs are discharges of raw or partially-treated effluent into waterways that occur when the flow through a combined sewer system (CSS) exceeds its capacity. CSSs are infrastructure common in older urban areas in the US constructed to carry stormwater and sanitary wastewater together through the same underground pipes. The EPA’s NPDES provides regulations and procedures for permitting, controlling, and mitigating the effects of CSOs. While NPDES mandates the elimination of CSO discharges during dry weather as a “minimum control,” dry weather discharges nonetheless can happen if the CSS is not functioning properly. More commonly, CSO discharges are prompted by heavy precipitation.

In 2004, it was estimated that 850 billion gallons of effluent is discharged annually from US CSO outfalls (see US EPA, 2004, for further background). Both CSOs and Sanitary Sewer Overflows, similar discharges from sanitary sewer systems, have recently been shown to lead to negative public health outcomes through an analysis of emergency room visits in Massachusetts (Jagai, DeFlorio-Barker, Lin, Hilborn, & Wade, 2017; Jagai et al., 2015). The public health hazard posed by these events is expected to increase as ongoing climate change increases the frequency and severity of extreme participation events (Patrz, Campbell-Lendrum, Holloway, & Foley, 2005; Patrz, Vavrus, Uejio, & McLellan, 2008).

4.3. Combined Sewer Overflow and Environmental Justice

We present an original analysis of the EJ impacts of combined sewage overflows, along with all data and code needed to reproduce the analysis, on the AMEND website (Sanders, 2018).

The EJ data used in this analysis comes from the US EPA EJSCREEN tool (US EPA, n.d.-a). Watershed and municipal-level EJ population characteristics are calculated by population-weighted averages over the Census block group-level data, with block groups assigned to watershed by comparison to geographic information system (GIS) shapefiles (MassGIS [Bureau of Geographic Information], 2017; US Census Bureau, 2017). Figure 2 shows the distribution of these characteristics across Massachusetts watersheds. The three urban watersheds of the Boston metropolitan area associated with the Boston Harbor cleanup (Dolin & Levy, 1990; Levy & Connor, 1992), the Charles, Mystic, and Neponset, have the highest levels of linguistic isolation as well as high levels of low income and non-white residents.

The CSO data used in this analysis was reported by the New England Center for Investigative Reporting (NECIR) based on their survey of New England CSO discharge reporting from calendar year 2011 (Struck, 2013). There have been substantial changes in population density, rainfall, and sewage infrastructure since 2011; however, more recent statewide or regional data is not available because there is not a standardized reporting system for these discharges (see Section 4.5) and the NECIR dataset is commonly cited (e.g., Wasser, 2019). The author explains:

“All states are required to regularly monitor bacterial levels in their waterways. But the EPA says it does not compile public records of where and how much sewage flows into those waters. Each state is supposed to report that information, but the NECIR inquiry found the data is often incomplete, inaccessible, sometimes handwritten and sometimes based on little more than guesswork, undermining the public accountability built into the Clean Water Act. (Struck, 2013)

The NECIR dataset, archived at the AMEND website, documents the source for each CSO outfall discharge estimate, which generally originate from “draft” state
Figure 2. Distribution of EJ characteristics across Massachusetts watershed populations.

Figure 3 shows the location of CSO outfalls in Massachusetts (points), with overlays showing the sum total of CSO discharge volume in 2011 by watershed. The interactive figure on the AMEND website will also display discharge volume by municipality and Census block group and allows viewers to zoom, pan, and to click on each CSO outfall point to view information about its location.

Figure 3. Interactive map of CSO locations and discharges per watershed and municipality.
discharge frequency, and volume. Similar maps on the AMEND website show the distribution of EJ communities across the state.

4.4. The Environmental Justice Consequences of Combined Sewer Overflow Discharges

We investigate the relationship between CSO discharge volumes and EJ population characteristics across all Massachusetts watersheds. We visualize the trend in CSO discharge volume by dividing the watersheds into four equal-sized bins according to each of the three EJ criteria defined in Section 4.1 and using bootstrap resampling to estimate the uncertainty in the population-weighted mean discharge volume estimate in each bin (Figure 4). We estimate the univariate dependence of CSO discharge on each EJ factor, and its 90% posterior (confidence) interval, with a simple population-weighted logarithmic regression model using a Bayesian methodology with weakly informative parameter priors (see e.g., Sanders & Lei, 2018) which is documented in detail on the AMEND website.

First, we explore the relationship between linguistic isolation and CSO discharge. The results suggest a statistically significant and high magnitude relationship between CSO discharge volumes and linguistic isolation. More linguistically isolated communities have much higher CSO discharge volumes on average. On average, watersheds that have twice the level of linguistic isolation tend to have 1.6 times (90% confidence interval 1.2 to 2.0 times) the level of CSO discharge (Figure 4a). The interactive version of this figure on the AMEND website has controls to show or hide the individual watershed points, which can be clicked to display detailed annotation.

Like the linguistic isolation trend, communities that are less predominantly white have much higher CSO discharge volumes on average (Figure 4b). We find that, on average, if a watershed has two times as high a concentration of non-white residents as another watershed, it will have 3.0 times (90% confidence interval 1.8 to 4.8 times) the level of CSO discharge.

Finally, Figure 4c shows the relationship between CSO discharge and income. Again, we find a strong and significant relationship. On average, when a watershed in Massachusetts has two times as many people in poverty as another, it tends to have 3.2 times (90% confidence interval 1.9 to 4.7 times) as much CSO discharge.

4.5. Dissemination and Impacts

We conclude that CSO discharges in Massachusetts substantially overburden contemporary EJ populations; the legacy of centuries of inequitable distribution of polluting infrastructure. Regardless of the historical factors responsible, our advocacy seeks for the Commonwealth to take action to resolve this disproportionate impact on its most vulnerable communities. While EJ has been a foundational principle of civic action around CSOs, to our knowledge, Section 4.4 provides the first publicly reported analysis of the distributional impact of CSO discharges on EJ populations.

In particular, we have advocated for legislation (Campbell & Provost, 2019; Jehlen, 2019) that would require timely public notification of CSO discharges, as well as reporting to an online statewide CSO database that could be integrated with AMEND. The introduction of such statewide monitoring, reporting, and notification of CSO discharges would enable residents to be aware of the public health risks generated when CSOs occur and would enable scholars and policy analysts to further study their impacts and make informed recommendations to mitigate their ill effects.

The findings of Section 4.4 were first presented to the Massachusetts legislature in June 2018 as part of a legislative briefing about the issue of public notification for CSO discharges in Massachusetts. In addition to these results, the briefing included presentations by representatives of Massachusetts civic organizations (the Massachusetts Rivers Alliance, MyRWA, and Merrimack River Watershed Council) providing context about the nature and history of the CSO issue and an overview of a predecessor CSO notification bill (Jehlen, 2018) under consideration that session. That bill was eventually passed by the Senate and referred to the committee on House Ways and Means, but was not voted on by the House. In the months since that lobbying effort, there

Figure 4. CSO discharge volumes versus EJ criteria across watersheds.
has been increasing public attention devoted to this issue in Eastern Massachusetts, highlighted by newspaper reports that cite the efforts of these water advocacy groups (Abel, 2018; Boston Herald, 2018; Eddings, 2018; Ottolini, 2018; Wasser, 2019). Most recently, the findings of Section 4.4 were submitted as written testimony to the Joint Committee on Environment, Natural Resources and Agriculture in April of 2019 as part of the first hearing for the bill in the current legislative session. This analysis and the AMEND web site are and will continue to be in use by our advocacy community throughout the legislative process surrounding this bill.

5. Theory of Communication

AMEND is located at the intersection of two prominent movements in the modern media sphere: data journalism and participatory journalism. The data journalism movement addresses how twenty-first century “data abundance, computational exploration, and algorithmic emphasis” (Lewis, 2015) manifest in the production and distribution of news (Coddington, 2015). The ability of individual advocates or small organizations to produce and publish reproducible policy analysis based on public domain datasets can motivate and support journalistic inquiry (e.g., Section 4.5) and connects specifically to participatory and communalist journalism, whereby individual actors can contribute to the gathering, synthesis, and dissemination of news and information (Kligler-Vilenchik, 2018; Ruotsalainen & Villi, 2018). Optimistically, open source platforms such as AMEND can help to address the problem of consolidation of knowledge about the manipulation and distribution of news and content relevant to civic engagement among an “information elite” with control of proprietary media outlets and massive networks of followers (Robinson & Wang, 2018) and the differentiated capacity between resource-rich and poor organizations to pursue data journalism (Fink & Anderson, 2015).

Open source data repositories make the tools of data gathering and analysis available to all individuals and organizations and lower the barrier to entry for their use. Platforms like AMEND can interact with other digital technologies like social media to enable political participation. Many studies have established that engagement in discussion on digital and social media is associated (through practice and perception) with increased civic engagement of various forms (Anderson, Toor, Rainie, & Smith, 2018; De Zúñiga, Jung, & Valenzuela, 2012; Obar, Zube, & Lampe, 2012; Saldaña & McGregor, 2015; Valenzuela, Kim, & Gil de Zuniga, 2012). Social media has even been suggested as a primary mediating mechanism, among digital technologies, by which civic engagement among individuals is transformed into political participation (De Zúñiga, Copeland, & Bimber, 2014). Moreover, the two-way exchange of information and communication through online public forums, including social media, fosters trust between institutions and their constituents (HarodeRosario, Sáez-Martín, & del Carmen Caba-Pérez, 2018; O’Connor, 2017; Warren, Sulaiman, & Jaafar, 2014).

By serving to increase the availability of public data resources and to enable policy analysis, platforms like AMEND generate another type of two-way communication complementary to the online social discourse: a kind of emergent data transparency cycle. For example, the case study in Section 4 illustrates how a civic actor (journalists) collects data (CSO discharge volumes) from public agencies (water infrastructure operators) that has not otherwise been published, that data is shared back to the general public through their reporting, then captured and integrated into a public data repository (AMEND), combined with other data published by public agencies (environmental justice population statistics) and enriched (through geographic analysis), shared back to the general public through reproducible online publication (on AMEND), and then used to advocate for the collection, preservation, and dissemination of additional data resources through the formal political process of state legislation. However, this example identifies a possible difference between data repositories and social media as a mechanism for political participation. Whereas the participatory impact of social media is often identified to be focused on action outside of formal political processes (Leyva, 2017; Theocharis & Quintelier, 2016; Vitak et al., 2011), public data repositories may generally rely on (or at least more directly link to) traditional institutions and formal political processes including interaction with data-publishing public agencies and the regulatory process governing their data transparency.

In this way, a web resource like AMEND can be thought of as part of the “textualization” process (Kavada, 2016) by which social movements can advance ideas, theories, and concerns raised during public debates, hearings, comment processes, and other interactions with government agencies (as well as observations of ecological conditions, public health outcomes, and other interactions with the natural and civic environment) into stable patterns of information that can be inspected, shared, and built upon. Because external pressure is a common driver of increased transparency and data publication among governments (Wang & Lo, 2016), there is reason to believe that this kind of cycle—with a loop of action culminating in legislative advocacy or other policymaking appeals—can successfully and sustainably iterate over time.

Ultimately, resources like AMEND that seek to increase access to public data and its importance in informing the public policymaking process serve to change the configuration of political agency (Kaun, Kyriakidou, & Uldam, 2016) in the states where they are deployed. By extending the decentralization of political discourse and action to information about policymaking contexts and outcomes and providing new forums for communication about this information, public data repositories and open source analysis platforms can play a role in a communications-oriented perspective on defining politi-
6. Conclusions

We have proposed criteria for online data resources that can help to build trust in policy analysis between civic organizations, agencies, and the public. We have presented AMEND, one such resource targeted for the Massachusetts environmental community that is designed around these principles, and presented a case study of the application of AMEND to the analysis of the impacts of CSO discharges on EJ communities. Finally, we contextualized this work and the AMEND resource in terms of concepts in media theory, suggesting that public data repositories be viewed as tools for reconfiguring political agency with connections to the movements of data and participatory journalism.

Going forward, we plan to introduce several additional enhancements to AMEND including:

- Additional data assets, such as Clean Water Act Section 303(d), impaired waters, assessment data from US EPA; data extracted from MS4 permit annual reports; and additional US Census data characterizing the municipalities within Massachusetts.
- Additional analysis articles, including analysis of the distribution of permit age by watershed and municipality and the effects of variation in budget and enforcement on 303(d) assessment outcomes.
- Improvements to the usability of the site to enable application of its data assets by less technical, more diverse stakeholders, especially through interactive plotting features to allow users to visualize interactive SQL queries through a web interface.

The modular design of AMEND is meant to facilitate portability to other contexts. Using our published code, other groups can launch their own version of this resource tailored for other communities or policy domains.

Acknowledgments

The author gratefully acknowledges Patrick Herron, Tracy Olson, and all the members of the Mystic River Watershed Association Policy Committee, the attendees of the Poliinformatics 2017 and the Boston Area Research Initiative Spring 2018 conferences, and the anonymous editor and reviewers for their helpful feedback throughout this work.

Conflict of Interests

No funding was provided by any source to support the development of AMEND. Sanders is a volunteer member of the Mystic River Watershed Association.

References


Coddington, M. (2015). Clarifying journalism’s quantita-


About the Author

Nathan Edward Sanders is a volunteer member of the Mystic River Watershed Association’s Policy Committee and is Chief Scientist for WarnerMedia Applied Analytics. Sanders also serves as the Chair of the national Leadership Team for ComSciCon, the Communicating Science Conference series. Sanders did his undergraduate work in Physics and Astrophysics at Michigan State University and earned his PhD in Astronomy and Astrophysics from Harvard University.