DEPARTMENT: SPOTLIGHT

Making Air (Quality) Visible

Exploiting New Technology to Dramatically Improve Atmospheric Monitoring

David B. Ramsay
MIT Media Laboratory

Joseph A. Paradiso
MIT Media Laboratory

Steven Hamburg
Environmental Defense Fund

There is an opportunity now to shape the future of air quality monitoring using the lessons from other industries. We advocate for an inclusive, open, distributed ecosystem that unifies regulatory-grade data with that of more inexpensive devices. This system incentivizes transparency and good data practices, encourages engagement, and accelerates learning. It creates the opportunity for sophisticated, distributed air quality models; supports community-driven innovation; and gets actionable data in the hands of those who care. This paradigm is not just the future for air quality, but also the future for sensor ecosystems in general.

Growing concern over air pollution has led to increases in air quality monitoring investment globally. In the US, more than 300 federal organizations collect air quality data, research funding has increased over the past several years, and citizen monitoring is on the rise.[1–3] With this increased focus, the overall market for air monitoring equipment is projected to grow from $3.4 billion in 2015 to at least $6 billion by 2022.[4]

Unfortunately, this mounting engagement has not resulted in a commensurate increase in our understanding of air quality or an effective market. The diverse backgrounds and incentives that have driven growth in the air quality measurement ecosystem have also prevented effective collaboration among its instrument manufacturers, application developers, research scientists, technologists, entrepreneurs, citizens, and policymakers. The status quo is failing to mature and deliver useful insights, despite the revolution in sensor technology.

We believe there is insufficient structure to catalyze a dynamic and rapidly evolving marketplace. The time is right for charting a vision forward as the private community grows, federal funding priorities shift, and affordable monitoring hardware becomes ubiquitous. An open and inclusive system based on communal norms would drive adoption, promote innovation, and

Published by the IEEE Computer Society
1536-1268/18/$33.00

ground interpretation of data in sound science. Implemented properly, this shift should benefit all parties across the value chain.

CURRENT STATE

Almost every pollutant can be measured using a multiplicity of methodologies and technologies, representing new opportunities and enormous challenges. The diversity of data from burgeoning technologies now on the market is subject to a wide range of operating conditions, detection limits, accuracies, and precision (both real and claimed). Once we include the range of techniques for processing these data, the resulting quagmire yields little of the promise of the sensor revolution. Even the best air quality monitors require routine calibration and human-in-the-loop quality assurance. The totality of these challenges is a susceptibility to error, even systematic error, which greatly complicates issues of data integrity and interpretation. In this landscape, an accurate, wide-coverage picture of the air we breathe is challenging to construct at best— for many devices, it may not be possible to extract useful information at all.

Compounding these technological challenges is the lack of a common ontology, even among knowledgeable and motivated parties. As a result, datasets are seldom combined, and the potential synergies are largely unrealized. Furthermore, institutional players often avoid sharing their data with the public (particularly highly time-resolved data) for fear of misinterpretation.

Unfortunately, the current generation of for-profit, consumer devices has an unreliable track record and is regarded with skepticism by the research community, despite the almost universal expectation that they will revolutionize our ability to monitor air quality.[5,6] Yet, portable devices are playing a large role in public discourse, saturating and diluting consumer interest with poor data quality and narrow actionability. In the worst cases, consumers are unwittingly making decisions based on unreliable or oversimplified information. Without a healthy ecosystem for sensor-derived data to be deployed in the service of society, the negative impacts of untrustworthy data will worsen as more citizens and communities fall victim to the unrealized hype of those peddling immature or poor-quality instruments and systems.

We believe there is an alternative path forward that counteracts the detrimental short-term market incentives for secrecy and exclusivity characteristic of proprietary development. Balkanization is avoidable by capitalizing on the synergistic effects of accelerated, aggregate learning— cooperation can be rewarded by society and the market. As low-cost monitoring hardware becomes more reliable and ubiquitous, an open-data system would allow researchers and app developers to explore statistical techniques that leverage higher spatial resolution and looser precision than is currently standard practice. The value for researchers, the for-profit sector, and civil society is indisputable—the challenge becomes how to steer the air quality community toward a more collaborative ecosystem.

As low-cost monitoring hardware becomes more reliable and ubiquitous, an open-data system would allow researchers and app developers to explore statistical techniques that leverage higher spatial resolution and looser precision than is currently standard practice.

LESSONS FROM OTHER INDUSTRIES

We can draw insight from other industrial sectors that have grappled with similar issues: the Internet of Things (IoT), open source (for example, Linux and Wikipedia) communities, and collaborative radiation monitoring.

An estimated 40 percent of the potential value for IoT applications is derived from interoperability, yet the relevant industrial players remain locked in a "standards war."[7] Large for-profits with the resources to build proprietary ecosystems have had an incentive to keep smaller players out. After decades of segmentation, the industry has only recently inflected toward possible standards consolidation. The IoT story is a cautionary one for emergent hardware ecosystems; the design of standards before the market demands it is a necessary requirement for rapid progress.

A positive counterexample comes from the open source movement: Wikipedia is successful because of the low transaction cost associated with contribution, its open license for the common good, and its transparent, self-policing nature (all modifications are recorded and made public).[8] Combining a low barrier for contribution with social good (common ownership and open licenses) has repeatably facilitated success and scalability in the open source world.[9] Wikipedia goes a step further with the use of automated “bots” for simple tasks like combating simple forms of vandalism, flagging high-risk data (potential “fake news” in current vernacular), and banning repeat offenders.

Perhaps the most instructive example comes from the Safecast radiation monitoring community that emerged after the 2011 Fukushima nuclear disaster in Japan. Like air pollution, radiation levels are complicated to measure, difficult to interpret, and political (affecting health, industry, and real estate). Despite these challenges, Safecast has succeeded in crowdsourcing the world’s largest open and reliable radiation dataset. Their success has been attributed to a focus on inclusivity and engagement combined with complete transparency in measurement methodology. Safecast has avoided restrictive data quality standards and has left explicit interpretation of their data to others.[10]

The threat of a fractured, rapidly growing air quality data ecosystem is real. However, it can be avoided by advancing an inclusive environment with low barriers to entry and clear social good. The private sector has a critical role to play in the emerging air quality ecosystem, but as seen in the examples above, only if the larger civil-society benefits are kept front and center through nonprofit engagement. While it is important to collect the information required to contextualize each air quality device, a highly centralized approach to data standardization, quality control, or interpretation is impractical and likely to fail.

> The private sector has a critical role to play in the emerging air quality ecosystem, but only if the larger civil-society benefits are kept front and center through nonprofit engagement.

Citizens will continue to collect and publish air quality data regardless of academic or institutional support. To compete with for-profits, the research community must be open with its data, inclusive and action-oriented with its technology, and clear about its interpretability. The goal must not be to decipher data for the community (an ever-moving epidemiological target); it must instead be to empower the community with the tools needed to comprehend the data for itself.

SYSTEM PROPOSAL

Interoperability and inclusivity are conflicting ideals. To satisfy both, we recommend a semantic web architecture similar to the World Wide Web. Semantic web technology preserves a minimal barrier to entry—as any data representation is supported—while advancing consolidation around robust, industry-driven ontologies and preserving data discoverability.[11] Semantic web principles are scalable, dynamic, and decentralized, and leverage existing web technology for data ownership and security.

An open framework invites data that vary from unusable to reference grade. Sensor metadata (device model, age, location, and so on) can facilitate numerous techniques for handling this variability. Simple approaches include reputation-based systems, while more complex possibilities that leverage known high-quality sensor data can automatically and algorithmically characterize new, co-located sensors.[12] We anticipate that data quality analysis will increasingly rely on contextual parameters from other data streams like traffic and weather, reinforcing the importance of a web-based ecosystem that helps blur the line between air quality data and other relevant information.

As more data are linked in this way, we expect rapid new insights as the successes of data mining and deep learning are brought to bear for environmental science. This shift will challenge the status quo methodologies of environmental research, and diminish the value of propriety algorithmic techniques. It represents an exciting frontier of distributed algorithm architecture, as web crawlers constantly search for useful training data over a vast network and update their models accordingly.

Furthermore, while organizations like the US Environmental Protection Agency (EPA) and the South Coast Air Quality Management District (SCAQMD) in Southern California independently test and validate portable affordable air quality devices, their results frequently stay unpublicized and their raw data remains internal, which is antithetical to building the trust and transparency required for rapid improvement. As we move toward open standards, it is our hope that these organizations and others that enter the testing space begin to open and cross-link their data, forming a collaborative to maximize impact and reduce unnecessary duplication. A robust and centralized collective testing infrastructure builds trust and transparency, ultimately shifting economic incentives toward preemptive, standardized, and open test datasets generated by the manufacturers themselves.

Data quality analysis will increasingly rely on contextual parameters from other data streams like traffic and weather, reinforcing the importance of a web-based ecosystem that helps blur the line between air quality data and other relevant information.

Efforts to create a shared data ecosystem are currently underway as part of the Environmental Defense Fund's Air Sensor Workgroup (https://www.edf.org/health/air-sensor-workgroup). We encourage interested parties to engage with and contribute to this growing effort. We hope that with the Air Sensor Workgroup and an Independent Testing Collaborative we can make progress on multiple fronts: (1) centralizing data so that it is easily discoverable; (2) standardizing data so that it is easily usable/sharable; (3) qualifying data with robust tools and mechanisms that ensure data quality; and (4) developing collaborative forums in which data can be meaningfully interpreted, shared, discussed, and criticized—perhaps an aggregation of the best practices from social media, Wikipedia, and emerging platforms such as PubPub.[13] As this effort grows, we need to track success by examining changes in cross-organization data sharing, emerging data standards, and raw data availability (especially independent, real-world testing data).

We encourage researchers to open their full datasets, document their measurement techniques, and make their processing algorithms open source. We encourage manufacturers to move toward transparent data capture and device validation in a way that builds community trust, rather than the all-too-common, black-box approach adopted by emerging private sector actors. Ultimately, we believe these recommendations can catalyze a thriving ecosystem that will reinforce the importance of these norms and standards. With support from the air-quality community, this ecosystem can be realized much more quickly and with the concomitant benefits to human health.

CONCLUSION

There is an opportunity now to shape the future of air quality monitoring using the lessons from other industries. We advocate for an inclusive, open, distributed ecosystem that unifies regulatory-grade data with that of more inexpensive devices. This system incentivizes transparency and good data practices, encourages engagement, and accelerates learning. It creates the opportunity for sophisticated, distributed air quality models; supports community-driven innovation; and gets actionable data in the hands of those who care. This paradigm is not just the future for air quality, but also the future for sensor ecosystems in general.

The time is now to create a symphony, not a cacophony, of air quality data.

REFERENCES

1. T. Dickinson, G. Paulson, and K. Sullivan, *Air Quality Observation Systems in the United States*, government report, Committee on Environment, Natural Resources, and Sustainability of the Nat'l Science and Technology Council, November 2013; https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/air_quality_obs_2013.pdf.
2. *Estimates of Funding for Various Research, Condition, and Disease Categories (RCDC)*, Nat'l Institutes of Health, US Dept. of Health and Human Services, 18 May 2018; report.nih.gov/categorical_spending.aspx.
3. A. Lewis and P. Edwards, "Validate Personal Air-Pollution Sensors," *Nature*, vol. 535, no. 7610, 2016, pp. 29–31.
4. "Air Quality Monitoring Market to Reach $6.0 Billion by 2022," P&S Market Research, 23 March 2018; https://globenewswire.com/news-release/2018/03/23/1449201/0/en/Air-Quality-Monitoring-Market-to-reach-6-0-Billion-by-2022-P-S-Market-Research.html.
5. E. Snyder et al., "The Changing Paradigm of Air Pollution Monitoring," *Environmental Science and Technology*, vol. 47, no. 20, 2013, pp. 11369–11377.
6. R. Williams et al., *Evaluation of Field-Deployed Low Cost PM Sensors*, government report EPA/600/R-14/464, US Environmental Protection Agency, December 2014; https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=297517.
7. S.L. Sharron and N.A. Tuckett, "The Internet of Things: Interoperability, Industry Standards & Related IP Licensing Approaches," *Socially Aware*, blog, Morrison Foerster, 2 February 2016; https://www.sociallyawareblog.com/2016/02/02/the-internet-of-things-evaluating-the-interplay-of-interoperability-industry-standards-and-related-ip-licensing-approaches/.
8. A. Ciffolilli, "Phantom Authority, Self-Selective Recruitment and Retention of Members in Virtual Communities: The Case of Wikipedia," *First Monday*, vol. 8, no. 12, 2003; http://uncommonculture.org/ojs/index.php/fm/article/view/1108/1028.
9. T. O'Reilly, "Lessons from Open-Source Software Development," *Comm. ACM*, vol. 42, no. 4, 1999, pp. 32–37.
10. A. Brown et al., "Safecast: Successful Citizen-science for Radiation Measurement and Communication after Fukushima," *J. Radiological Protection*, vol. 36, no. 2, 2016, pp. S82–S101.
11. T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities," *Scientific American*, vol. 284, no. 5, 2001, pp. 34–43.
12. D.B. Ramsay, *LearnAir: Toward Intelligent, Personal Air Quality Monitoring*, thesis, MIT Media Lab, August 2016; https://resenv.media.mit.edu/pubs/theses/david_MS.pdf.
13. T. Rich, "Collaborative Scientific Publishing: A New Research Ecosystem," thesis, MIT Media Lab, May 2017; https://dam-prod.media.mit.edu/x/2017/05/03/trich-phd-17_AZO805s.pdf.

ABOUT THE AUTHORS

David B. Ramsay is a PhD student in the Responsive Environments Group at the MIT Media Laboratory. Contact him at dramsay@media.mit.edu.

Joseph A. Paradiso is an Alexander W. Dreyfoos (1954) Professor in Media Arts and Sciences and director of the Responsive Environments Group at the MIT Media Laboratory. Contact him at joep@media.mit.edu.

Steven Hamburg is chief scientist at the Environmental Defense Fund. Contact him at shamburg@edf.org.