HEALTH & ENVIRONMENTAL FUNDERS NETWORK

Building Power through Data-Sharing:

Issues and Opportunities for Environmental Health and Justice Funders

David Eaves - August 2014

Executive Summary

This paper is intended to provide environmental health and justice funders with background information that may be helpful in considering investments to increase data-sharing. It notes relevant trends in open data and technology, highlights examples of data-sharing within the environmental health and justice movement, and provides lessons for funders.

The paper was produced for a June 2014 meeting of environmental health and justice funders and therefore focuses on examples within that field. Many of the issues outlined also may have broader applicability for funding data-sharing in other social movements.

Introduction

In 2007, during a divisive and deeply contested Kenyan presidential election, a group of software developers created a platform that allowed citizens to report acts of violence via email and text message. The new platform Ushahidi – meaning "testimony" in Swahili -- had a profound impact on media coverage of the election by increasing reporting on the violence. Ushahidi has since matured into a widely used platform for mapping crises and problems.

This was not the first time people attempted to crowdsource a problem; Wikipedia, for example, was founded in 2001 as an online encyclopedia that anyone can edit. However, Ushahidi seemed to capture the public's and the technology community's imagination in a profound way, prompting interest in how large numbers of individuals could share information to confront injustice.

This paper explores how trends around data, and in particularly the *sharing* of data, could similarly impact and be leveraged by people working for environmental health and justice. The paper will discuss:

- Lessons and trends from the broader open data community that could be relevant for the environmental health and justice field;
- Insights about current environmental health and justice data-sharing efforts; and
- Key issues, opportunities and questions related to data-sharing for environmental health and justice funders to consider in grantmaking.

Terms Defined:

Data: Quantifiable pieces of information or facts, e.g. a measurement or location. A written document or analysis would not constitute "data."

Crowdsourcing: A process of obtaining input for a task by enlisting the participation of a number of people, either paid or unpaid, typically via the Internet.

This paper will not dive deep into the importance of data more generally. While data is clearly going to become more and more important to the environmental health and justice sector (and is relevant to the context to this discussion), this paper will focus on ways the *sharing* of data could help the movement's many actors advance their interests.

Broader Context: Data and Collapsing Costs

Today, we live in a world obsessed, intrigued and concerned by data. Big data, open data, personal data, raw data, data mining are all terms that 10 years ago would have been foreign to most people and yet today are common parlance. This explosive growth in the amount of data being collected - as well as our growing awareness of its importance - is driven by two factors:

- **Costs of collecting data keep declining.** As the cost of computers and hardware continue to collapse, it has become easier and easier to collect more data about anything, from pipelines performance to asthma inhalations to email usage.
- **The internet's growing ubiquity.** As more and more of what we do uses an internet-connected device, tracking and gathering data can become a simple by-product of these activities, rather than something explicitly designed.

So, while this paper is not about data, it is worth noting its growing role in shaping and driving public opinion, as illustrated by my own experience as a negotiation advisor with the Canadian Boreal Forest Agreement (CBFA).

The CBFA is a landmark agreement between many of Canada's largest forestry companies and environmental non-governmental organizations (ENGOs), aiming to create a regime for sustainable wood harvesting in Canada's Boreal Forest. To implement this on a regional basis, data is "mashed up" from maps provided by ENGOs, the companies and public government maps. The goal is to use maps with woodland caribou (a marker species) roaming areas, critical landscape features, historical logging and other data, to find places to stop, defer or accelerating logging in a way that preserves local economies and forests.

While such a negotiation, at a principle level, would have been possible 20 years ago, the detailed maps create new obstacles (one can see how bad things really are) and new opportunities (granularity provides more flexibility in finding specific areas that can and cannot be logged).

Similarly, environmental health proponents are drawing data and maps together in new ways to build knowledge and accelerate action. Examples include California's CalEnviroScreen identifying neighborhood-level conditions drawing together data on environmental hazards, health impacts, and demographics; or FracTracker, an open data platform enabling mashups of governmental, industry, academic and citizen data on hydraulic fracturing activity.

Terms Defined:

Data Mash-Ups: Combining two or more sets of data to produce new content, such as a map showing facilities, permits and violations along with information about the surrounding area.

Such maps and data are changing expectations of what is possible, the ability to measure impacts, and the skills needed on all sides (industry, civil society, government) to participate.

The perceived growing importance of data has thus created growing demand for the "supply" of public interest data. Some of this has happened in the environmental sector but nowhere has the push for "open" data been greater than in the government sector.

Broader Context: The Open Data Movement

In February 2009, Sir Tim Berners Lee (inventor of the World Wide Web) pleaded in a TED talk with organizations -- and governments in particular -- to make data they collect about virtually everything open: free for anyone to download, remix, make use of, and share. Since then, there has been a veritable explosion in the amount of data governments share openly with the public.

Of course, governments have been collecting and sharing environmental, weather, tidal and other data for hundreds of years. Such data has been seen as essential for commerce, used by sailors, insurers, the military, and others to plot and plan trading routes.

In the 1980s, the US Environmental Protection Agency (EPA) began to release a Toxic Release Inventory (TRI) of data on polluting emissions from facilities across the country. The act of simply creating the TRI is believed to have had a salutary impact on pollution emissions, as factories sought

817 Silver Spring Avenue, Suite 110 • Silver Spring, MD 20910 • Tel: (301) 565-0500 • E-mail: info@hefn.org • www.HEFN.org

to improve their standing within the database, as well as to diminish negative local reaction to chemicals emitted in communities and adverse investor reactions to risks created by toxic emissions.

The open data movement of the 21st century has had multiple goals. Proponents within government often see open data as a way to drive innovations, find opportunities to improve efficiency and competitiveness, and create jobs; this is also a vocal group of activists who see open data as a way to foster transparency and accountability in government.

Among important successes to date, the open data movement has provided a growing, data-literate community of social activists with new information for advocacy and public policy discussions.

In my own city of Vancouver, BC, one of the most voracious consumers of local government open data is a local think tank run by an architecture firm. Their highly-acclaimed work on the threats to property tax and city life posed by rising sea levels was made possible by downloading open local and state government data on building footprint, property assessment and topography. Similarly, the Open Data Institute incubator company discovered potential savings of hundreds of millions of dollars through analysis of generic versus branded drug use across the United Kingdom.

These examples, where locally driven data activists and emerging companies find ways to create social benefit and save money using public data, continue to drive the belief that by making data open, new public and private benefits can be created.

For funders considering investments in open data sharing, some important lessons are worth noting from government open data efforts:

- **Costs**: Collecting and sharing even "free" data is not cost-free. Time and resources are required to make it available and usable.
- **Demand**: Demand will not always magically appear for data that is released.
- **Capacity**: Many, many sectors of society remain functionally data illiterate; their ability to consume, work with and make use of data remains very limited.
- **Equity**: Open data-sharing can help democratize knowledge or – in light of cost and capacity issues – widen inequities.
- **Standards matter**: Data becomes significantly more powerful when it can be compared across organizations and geographies; standardizing data allows it to scale.

Why Sharing Matters

For many, the promise of open data is that it will let more and more people leverage a common resource to solve an array of problems. Understanding *why* data-sharing can help make this happen is essential -- and this is commonly misunderstood. A model for understanding what it can and cannot do can be found in open-source communities.

Open-source communities – often made up of dispersed and unaffiliated or loosely-affiliated contributors – have developed some of the most widely used software in the world. Why are these communities effective in marshalling such dispersed (and sometimes vast) contributions into powerful programs? And why do they enable co-creation?

Lesson #1: Effective open source projects understand and find ways to advance the interests of their contributors.

One common misconception is that a vast army of tech-savvy volunteers awaits and that, by simply making data, information or software code available, this crowd of volunteers will flock to the cause and help increase a movement's power and capacity.

Much like the tech company that goes on to become Google, this narrative exists... for about 0.0001% of collaborative projects. For everyone else, the reality is far more challenging.

Most commonly, people will focus on (and volunteer for) tasks that advance their interests. Said interests vary; they could involve destroying a monopoly (Firefox in its heyday), exposing violence and injustice (Ushahidi in Kenya), having fun and gaining minor fame (a number of Call Me Maybe YouTube videos), building a new skill (many open source projects), making money (Drupal developers) or simply being part of a grand project (Wikipedia editors).

817 Silver Spring Avenue, Suite 110 • Silver Spring, MD 20910 • Tel: (301) 565-0500 • E-mail: info@hefn.org • www.HEFN.org

In the world of government data, I mostly see people downloading and using government data for the following reasons:

- To advance a business interest (e.g. better understand a market place or competitor);
- To engage in a policy problem (e.g. pipeline spill, school attendance or bike accident data); or
- To learn (e.g. wanting to learn how to make visualizations or understand better how government works or for writing academic or non-academic papers).

Most open government data portals have an exceptionally small number of users compared to the populations they serve. In the end only a tiny fraction of people directly engage with government data. This reality should be a cautionary tale for those who believe that simply making environmental health data more available will, in of itself, create a large (or even modest sized) community of users.

The importance of understanding people's motives becomes even stronger when one looks not just at people making use of data but contributing data. **It is comparatively easy to set up the infrastructure to accept data. The real challenge is cultivating a community of contributors willing to spend the time and energy to contribute information.** For many potential contributors, their willingness to contribute strongly depends upon the impact they feel their contribution has.

For example, projects like "I paid a bribe" in India, which lets contributors track where and why they paid bribes, has seen contributions drop off as users question whether the project has altered the practice or even precipitated much debate. And there are many a "dead" Ushahidi map in which contributors never materialized, often because gathering the data was important to the organizer of the project but offered no tangible benefit (in awareness or systems change) to potential contributors.

Lesson #2: Cooperative systems may scale easier than collaborative ones. It is important to recognize how much "friction" or communication cost a sharing or co-created initiative will require. The co-creation processes I've observed that tend to scale (a meaningful though imperfect proxy for success) do so *not* because they make people work together more. Quite the opposite! Their success often springs from their ability to foster sustainable systems that are, in many ways, anti-collaborative.

Terms Defined:

Source Code: The lines of text (sometimes including comments or notes) that constitute a software program.

Open-Source: Originally refers to a computer program whose source code is freely available to anyone. Projects which similarly share basic intellectual content may also be referred to as "open source."

Open-source community: A group involved in developing or using some open source software, such as the popular Firefox internet browser. May also refer to groups advocating more open sharing of source code and data.

To clarify, let's define two types of ways people may work together: collaborative and cooperative.

Collaborative processes enable differing and possibly conflicting views to merge and create something new and previously unimagined (think of Hegel's thesis and antithesis coming together in synthesis). Collaboration is high-touch. It involves people working together tightly and regularly.

Cooperative processes, in contrast, allow people to leverage one another's work, but they may require minimal, if any interaction.

Many online projects referred to as collaborative are, in reality, cooperative. They break problems down into discrete smaller pieces. Dividing up a big problem makes it easier for a network of volunteers to opt-in and provide solutions in an efficient manner. However, this efficiency is created in large part because those involved in the project *don't* need to talk to, exchange ideas or even interact with one another.

This is precisely the genius of open source projects. They architect problems into small enough "chunks" that an individual may grab one, go off and *work in isolation*, bring back their proposed solution and easily integrate it back into the whole. This is not to say that *all* problems (or every step of the process) in an open source community are cooperative.

817 Silver Spring Avenue, Suite 110 • Silver Spring, MD 20910 • Tel: (301) 565-0500 • E-mail: info@hefn.org • www.HEFN.org

But the more that problems (and steps) can be converted from collaborative ones to cooperative ones, the more the community can scale and the faster it can do more and more work.

Herein lies the opportunity for open data. The goal is not to force people to work together, but rather to create a permission free environment where people can either:

- Take a data set and go off and use it to try to solve a problem they find pressing; and/or
- Contribute data to a dataset and thus make it more effective for them (and others) to use it to solve a problem they care about.

This is not to say that open data can't be used by groups engaged in high-friction collaborative processes, working closely together to achieve goals they could not on their own. But this is not necessarily its main value. Rather open data is about enabling anyone to use and/or contribute.

A few important points for funders considering data-sharing investments emerge from the above analysis:

- Sharing data is about creating a shared asset that may be valuable to a larger community.
- It is essential to understand the motivations of people you hope will contribute or use data. If they have no incentives to do so (e.g. if it does not advance their goals), they will not share or use it, no matter how effective the tools are.
- Encouraging data-sharing is a terrible way to *compel* or even just encourage disparate groups to collaborate.
- Encouraging data-sharing can be a helpful way to enable disparate groups to cooperate in expanding a community's shared assets.

The Common Mistake: Focusing on the How

One of the most common errors I see when institutions consider how data-sharing could help their cause is to focus on the technology and tools. The reasons are understandable. The costs to sharing data are lower than ever. Creating an app that crowdsources data, or a website to share data has, by almost every measure, never been easier. The desire to simply pick a solution and run with it is tempting.

The problem, as highlighted above, is that the real costs are not in setting up technology solutions for sharing data. It is in what surrounds the tool: finding ways to cultivate a community or users or contributors; keeping the data up to date; possibly educating people about how to interpret or understand that data. **Most of the very real costs to sharing and co-creating data have nothing to do with the technology.**

Governments may be able to simply dive in; their data is a public asset and they may be less resource constrained. But nonprofits with limited resources need a strategy. Why are they sharing? What problem is it solving, and for whom? Will it save them money? Advance their interests? Generate power for their organization, for other stakeholders, for the movement?

The Key: Focusing on Why

As outlined above, collaboration and to a lesser degree cooperation each carry costs. Before any organization embarks on data-sharing, it's important to know why it seeks to do so, and to evaluate whether the outcomes of sharing or co-creating will be more valuable than their costs. **For non-profits and funders, an even more important question is how sharing would help create power for an organization, for other stakeholders, or for the movement.**

While increased sharing of data can have a number of benefits, my interviews with several data-sharing projects being funded by HEFN members identified three broad categories of common problems in this field for which data-sharing seems to be particularly effective at solving problems and/or creating power.

These categories of problems are not perfectly delineated, and there is overlap between them, but each highlights some issues that may be helpful to funders considering investments in data-sharing.

Category #1: Data Organizing Problems

Some environmental health and justice efforts focus on problems where there is a lot of information but it may be dispersed, contradictory, duplicative, and/or have significant gaps. A major challenge is harmonizing or "standardizing" data so that different bits of information can be readily compared. By organizing previously incoherent or hard to locate information, the new datasets can be used to pressure stakeholders, inform

817 Silver Spring Avenue, Suite 110 • Silver Spring, MD 20910 • Tel: (301) 565-0500 • E-mail: info@hefn.org • www.HEFN.org

communities or markets, and thus build power and change behavior.

A perfect example of this is the Health Products Declaration Collaborative (HPD). This collaborative, launched with support of building product manufacturers and the Healthy Building Network's Pharos Project, has focused on creating a data standard for reporting building product content and associated health (and hazardous impact) information.

Markets often react to more data. By stimulating demand for this data among architects and buyers, the HPD Collaborative is creating a data set that can help bring greater transparency to the building marketplace. The resulting data set will better enable discerning buyers to connect with sellers, encouraging a market segment that favors less toxic building materials. The very existence of the dataset itself serves to educate buyers. It builds awareness that toxicity is something they *should* care about, and gives them an actionable place to explore materials through this lens (as opposed to exclusively by cost).

Moreover, there are powerful reinforcing incentives at work here. Building materials buyers seeking products with lower health impact have clear motivations for using the data set. Sellers seeking to appeal to those buyers have good reason to contribute accurate, up-to-date data in the system, since their value proposition to the healthy-materials market segment is linked to the chain of custody and degree of toxicity of their product.

Data-sharing about product content and health impacts also pressures other market actors to rethink product design and manufacturer. Absence from the list not only removes sellers from a market segment but is potentially brand-damaging, as buyers may wonder why a company is unwilling to disclose toxics information about its products.

With data organizing problems such as building materials and their health impact, finding such incentives to encourage data-sharing is sometimes important. It's also worth noting that incentives to share may not always be the main challenge. Sometimes potential participants may already be sharing this data (say on their website).

The challenge is getting everyone to share the data *in the same way* so that it can easily be compared and compiled. The hard work thus usually involves creating a standard and persuading potential participants to adopt it.

This is precisely what the HPD Collaborative described as the main obstacle to their success. The project had to focus on persuading the relevant stakeholders to take an interest, and on getting them to agree on a common standard for sharing data. For data-organizing problems like these, much of the significant investment is upfront. For instance, standards development is a highly collaborative and often costly process but one that, if successful, can then enable tremendous scaling-up.

This is what makes data-sharing to address data organizing problems so exciting. In the case of the HPD Collaborative, with a data standard in place, any company in the world may start complying and sharing the relevant information about its product, at a relatively low cost to both the collaborative and the company.

Category #2: Data Identification Problems

For some environmental health and justice problems, the data sought is hard to gather. For example, where are the environmental hazards in a community, which are priorities for community action, and who can address them? For such situations, the challenge may be less about organizing available data (although this may be a goal) as about a more immediate goal of trying to collect it in the first place.

Unlike data organizing problems -- which are often the domain of geeks and insiders (how many people have tried to create a data standard?) -- data identification problems can often involve the general public and may use methodologies that have most captured public interest in technology and social change.

The previously described story of Ushahidi – a tool initially created to enable people to report incidents of violence in Kenya, which evolved into an all-purpose mapping platform for reporting any issue -- is a perfect example of an identification problem.

Among many examples of data being gathered and shared to identify environmental health and justice problems, the work of the Comite Civico Del Valle (CCDV) in Imperial County, California comes to mind. CCDV was initially founded in the 1980s to promote the voting rights of migrant workers and to improve access to education for their children.

817 Silver Spring Avenue, Suite 110 • Silver Spring, MD 20910 • Tel: (301) 565-0500 • E-mail: info@hefn.org • www.HEFN.org

Around 2000, with large-scale burns in farming fields and local industry's toxic emissions negatively impacting the health of the community, the organization began to address environmental health issues.

At the suggestion of a young California Environmental Protection Agency (CEPA) employee -- and with a special grant -- the CCDV built a Ushahidi web platform. It began to gather data about environmental health hazards from community residents and to map those incident reports.

Having its own environmental monitoring data has proved critical to the CCDV. Previously, community groups reporting problems to state or local authorities experienced a "black hole" as problems disappeared into a bureaucracy where they were hard to trace or to consider in context of other incidents. With its own data about environmental incidents within the valley, it became easier to spot trends, problems and repeat offenders.

This in turn helped the CCDV build power. The CEPA was both impressed and concerned with the CCDV's information and so tasked one of its officers to engage with the CCDV directly, enhancing CCDV's ability to get the attention of local officials and responsible parties. The map also made it easier for the CCDV to get further funding as it could better articulate and visibly show its work and the results generated.

Above all, creating an open data set to compare to those of the regional water quality control and air pollution boards allowed the CCDV to confront local authorities often co-opted by local industry and/or farm owning groups. Previously, it was easier for local authorities to ignore incidents as they were the only ones who knew of the scope of the problem. The transparency created by the CCDV map made it harder to ignore community hazards.

Indeed, so powerful was the CCDV's map that it attracted pushback. The CCDV's incident reporters were labelled vigilantes. The group was attacked for lacking expertise. The data was attacked for not being reliable enough. Such efforts by local power-holders to undermine the credibility of the data is an indication of how powerful it was.

Interestingly, state and federal agencies were significant supporters of the CCDV's maps, providing eyes and ears on the ground that state and federal officials could not hope to replicate.

It is worth noting that, unlike data-organizing or science problems, the accuracy of the data in identification problems is actually not as important. This type of approach is much more tolerant of errors. This is because actions (like an enforcement action) rarely depend on any single data set. Making it easy for many people to contribute data is often more important than accuracy. So long as the combined data is broadly accurate, it can create a meaningful context for highlighting issues, fostering engagement, and creating power.

Today the CCDV's mapping platform is known as the Innovation Value Access Network or IVAN - and it has spread to several regions in California. More intriguingly - and exceptionally rare for these types of crowd-sourced approaches - the CCDV is working on ways to connect its database with the official database at various environmental agencies. This would allow problems reported on IVAN to be inputted directly into the authorities' systems as well.

The danger with the CCDV's story is to be lured into the simple narrative. Local group creates a map. Given a new tool, local people rise up and report problems; this collective action creates leverage that is then used to bring about change. On one level this narrative is true, but it risks making the technology – the Ushahidi platform -- the magic ingredient.

The truth, as Luis Olmedo of the CCDV likes to point out, is that there is "a huge human dimension" when talking about technology and sharing information. Many of the things traditionally required to make a social movement -- building a community of people willing and able to contribute data, figuring out how to use that data to pressure people, cultivating allies, and ultimately bringing people together to solve problems -- are still required. Co-creating data is not a replacement so much as a force multiplier.

For those looking for other examples of the use of data-sharing for identification problems in the environmental health space, Risha Bera and Anna Hrybrk's paper iWitness pollution map: crowdsourcing petrochemical accident research is an excellent overview of how a community in Louisiana used Ushahidi to report Oil Refinery accidents and mobilize local resources to their cause.

817 Silver Spring Avenue, Suite 110 • Silver Spring, MD 20910 • Tel: (301) 565-0500 • E-mail: info@hefn.org • www.HEFN.org

Category #3: Scientific Problems

Scientific problems are those in which data is collected or shared in order to quantify or understand the nature of something. One distinction between scientific problems versus identification or organizing problems is that the quality of the data required is usually very high. Academics, particularly in the hard sciences, usually need to know where data came from and the methodology behind its collection. This usually creates barriers to the sharing of data. It can also often create barriers to its use, since possessing the knowledge and skills to understand these issues can be a precondition to use.

Another distinction is that efforts that focus on scientific problems are often trying to help foster a community of knowledge, using data to bring together people with expertise who might find each other's research interesting and therefore advance the problem.

The purpose of scientific problems is usually not to mobilize a community or communicate with the masses, but to use data as a way to bring together expertise to more fully understand a problem and/or to develop inputs for policy responses.

The Endocrine Disruption Exchange (TEDX), for example, has created a database of studies and data related to chemical exposures and environmental health. One goal of the site is to overcome disciplinary boundaries to further endocrine disruption research.

According to Sara Wylie, Northeastern University Assistant Professor of Sociology/Anthropology and Health Sciences, it has been "a successful model of connecting researchers working on related issues in different scientific fields, with non-profit organizations and policy makers."

Another good example of a project addressing scientific problems is DataBasin. Managed by the Conservation Biology Institute, this environmental mapping non-profit has become part geospatial consulting and analysis shop, part data education and training organization, and part data portal for researchers and policy makers.

For both DataBasin and TEDX, the data is as much bait as it is product. The goal, at least in part, is to attract key stakeholders with valuable skill sets, with a bigger goal generally of fostering a knowledge community. For these projects, community management skills are critical to success. For DataBasin, an entire community section of the website is populated by over 200 groups that have organically developed around a shared interest, whether geography, species or policy.

As DataBasin researcher Tosha Comendant pointed out: "The groups are a real success. For example, an endangered habitat group in San Diego used land parcel information from DataBasin for a proposal that was funded that allowed them to acquire critical land. We see a lot of that: 'DataBasin helped the proposal get funded'."

For funders considering investments in data-sharing, considering the type of problem a project aims to solve – identification, organizing, scientific – may be helpful in reviewing the level of investment, data quality, stakeholder engagement, and skillsets required for success or scale.

Other Trends

A few other key trends are worth considering as funders look at data-sharing.

Lower coordination costs
A growing number of tools make it easier and easier for people to coordinate the sharing of data, information and other assets. Projects like Mapmill.org make it easier to crowdsource quality images to make up large area maps. Tools like Flickr make it easier for people to share photos of oil spills or other visually-notable incidents. In Toronto, I've been working with the non-profit Acorn and the Toronto Star (a newspaper) to use SeeClickFix (a problem mapping tool) to track deficiencies in public housing.

The growing array of such tools devised for data-sharing and organizing, as well as a growing number of free or low cost commercial tools (Flickr, SeeClickFix, Ushahidi), may be repurposed for activist and scientific ends.

Such tools for creating and sharing data will continue to get cheaper. However, the costs of organizing volunteers, managing a community and *figuring out what type of information will get you power* will not decline as quickly, if at all. Investments in this tool-surrounding capacity are a continued need and philanthropic opportunity.

Sensors
In addition to more tools for sharing data, the ways of *gathering* data are also getting cheaper.

817 Silver Spring Avenue, Suite 110 • Silver Spring, MD 20910 • Tel: (301) 565-0500 • E-mail: info@hefn.org • www.HEFN.org

Numerous environmental justice groups have used "bucket brigade" technology using simple techniques to take air pollution samples. An Air Quality Egg, another tool that allows any household to measure the air quality outside its home, can be bought for $150.

Costs of such citizen data collection are sure to decline further. Every smart phone now comes with a camera. When connected to kites and balloons, these phones suddenly became mapping tools, as demonstrated by projects like Public Lab which facilitated citizen photography and collective photo-mapping of the Deepwater Horizon oil spill. Likewise, NoiseTube converts the microphone of any smartphone into a noise pollution measuring and mapping device.

As hardware costs continue to decline, and more and more devices can be connected to the internet, the costs of sensors, and thus the ability to gather more and more data will explode.

Open Government Data
The United States is already a leader in sharing government data. Trends towards more data transparency, such as through the Obama Administration's national action plan on open government, may make public more federal and state government data sets relevant to the environment and health. This would give many stakeholders, including industry, academics and nonprofits, access to potentially valuable information.

Many environmental health and justice groups already find public data of high value. For instance, Luis Olmedo of the CCDV talked about how the EPA's Toxic Release Inventory helped his stakeholders focus on some likely targets for illegal or problematic emissions. Indeed, CCDV was able to track toxic releases from facilities that had failed to disclose any releases to the EPA in over a decade.

One area of potential action for the environmental health movement is to advocate for more collection, public release and improvement of government data. This could, for relatively low cost, leverage access to large quantities of high value data that could prove particularly critical for scientific problems, as well as for some data identification and organizing problems.

Data-Sharing: Lessons for Funders

Lesson # 1: Data-sharing/gathering can be low-cost, but it is not cost free.

There is a risk to believing that the software and hardware costs of sharing data have declined so much that sharing data is a simple, low-cost move. While technology costs may decline, shifting to sharing data usually carries with it organizational and capacity challenges, particularly for under-resourced non-profits.

This is not to say that organizations should avoid sharing data. In fact, more data-sharing should generally be encouraged. However, funders should be aware that this will entail real and, in some cases significant, costs even when technology is free or cheap.

Lesson #2: Have a goal for sharing data -- and focus on building power.

Be clear about why you want more data-sharing. Is sharing the data critical to the mission, a by-product of other activities, or just something the organization feels it should do? The particular reason an organization has for sharing data is far less important than its *having a well-considered reason*. This reason should guide the resources it puts into the problem and provide it with ways of measuring success.

If data-sharing is critical to success, it is essential that the organization have not only the technological capacity but also the soft skills and processes required to, for example, foster a community of contributors, or to build an alliance around data standards. Data-sharing problems rarely end or revolve around technology choices; getting everything else right is equally and often more important.

Ideally, any goal for sharing data should revolve around how it builds the organization's power, and/or that of the movement as a whole.

Lesson #3: There are already data-sharing success stories within the environmental health and justice movement.

Examples cited in this paper highlight ways in which environmental health and justice groups have been making effective use of data-sharing. Interviews conducted for this paper focused, admittedly, on some organizations on the leading edge of the field's data-sharing efforts; the stories shared therefore may not be representative of the capacity across all HEFN members' grantees.

817 Silver Spring Avenue, Suite 110 • Silver Spring, MD 20910 • Tel: (301) 565-0500 • E-mail: info@hefn.org • www.HEFN.org

However, the range of examples cited does illustrate a diversity of ways in which data-sharing is already advancing environmental health and justice interests. It also flags groups within the field that could help mentor others about how to replicate successes. The field's early adopters -- both those with successes and failures under their belt -- are an important resource for HEFN and should be leveraged for webinars, best practices and other mechanisms helping others learn and adopt these practices.

Lesson #4: Fund tools that serve the goal.

Funders and their field partners should think about what capacity each individual organization needs to build, versus those that could more efficiently be centralized in a single, more focused entity and leveraged by others. DataBasin is an interesting model for exactly this type of work. It specializes in hosting data for a number of other entities, sparing others the need of developing this capacity. DataBasin has also moved up the value chain, offering geospatial analysis and consulting services, a capacity that many smaller non-profits would struggle to cultivate and might distract them from their core mission.

Another advantage of centralizing some of the capacity is that it can force a movement to adopt some common standards around the collection and sharing of data. This could make it easier for data sets to be mashed up and compared in the future, revealing correlation and potentially new insights.

Lesson #5: Data literacy will become an increasingly important capacity.

As data becomes more accessible, it will be important for environmental health and justice groups to be data literate. For some, this will mean being able to analyze data and use it to draft policy or respond to a claim. For others it may mean creatively using data as a tool in a campaign or to shift a market place. It may also mean being discerning about how to use data to advance an organization's interest, or being conscious of how data might be used by others (industry or government) to compromise goals.

A few steps funders could take are to become more aware themselves of skills that are needed. Much like one might be cautious of funding a project proposal built around community funding from a non-profit with no experience in community organizing, funders need to be aware of when data skills are needed for a project to be successful, ensuring that the organization has the capacity to meet the project's goals. As sharing data on a large scale is a relatively new phenomenon, the opportunity for experimentation is great, but so too is the opportunity for organizations to be funded to do work for which they are ill-equipped.

Several organizations, such as the aforementioned DataBasin, provide training around GIS tools as well as more general "using data" courses that could be helpful to environmental health and justice funders and/or advocates. It might also be worthwhile for HEFN to host an annual roundtable of data-sharing projects, so that best practices and learnings can be shared more quickly.

For example UnionSquare - a venture capital company that is a private sector equivalent of HEFN -- regularly brings together people in similar roles across the various start-ups it funds. This allows them to learn from one another more quickly. If, as I suspect, the use and more importantly, the sharing of data is to become more common across environmental health and justice grantees, finding ways to scale and increasing data literacy will be worth funding in its own right.

It might also be worth sending some representatives to the Open Source Community Managers' conference that takes place before Open Source Convention in Portland, Oregon every July. While not a perfect fit, this is one of the largest meetings of people who think about how to build online communities around the world.

Lesson #6: Other skills will still matter in a world of data.

Many data collection problems are, as Luis Olmedo put it "human problems." Invariably the skills of negotiation and community building are essential to either creating or sharing data effectively. Funders could encourage organizations intending to build themselves around a strategy of data-sharing to not focus exclusively on the technology, but to have the full range of capacity necessary to cultivate a community of contributors or users.

As noted above, funders should make sure that grant applications are not too narrowly construed. Projects that involve data but exclusively focus on organizing people and ignore technology considerations are as dangerous as projects that fetishize technology and fail to consider how to organize and engage the people involved in the problem. Thus it is critical that funders *ask the right*

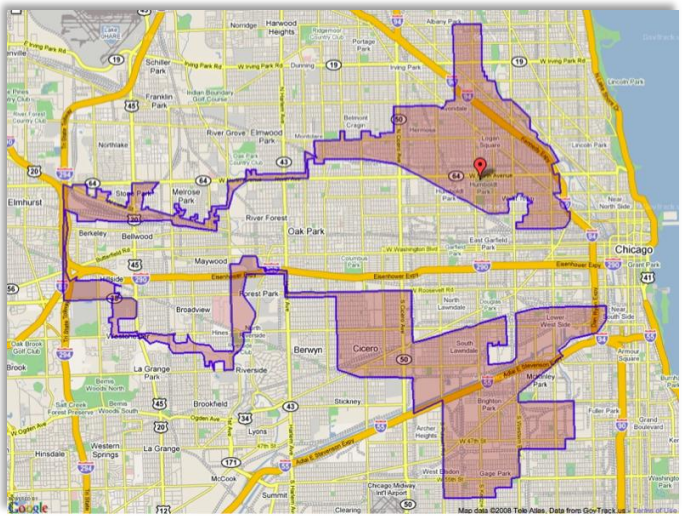
817 Silver Spring Avenue, Suite 110 • Silver Spring, MD 20910 • Tel: (301) 565-0500 • E-mail: info@hefn.org • www.HEFN.org

questions and ensure that a *broad understanding of capabilities* is considered in grant applications.

Lesson #7: Data is a not a panacea.

Michael Bloomberg, the former Mayor of New York, is famous for his application of data to solve many of New York's problems. Indeed there are several amazing examples of the city saving millions of dollars and people's lives by carefully examining data and trends and refocusing resources. This approach is nicely summed up by the mantra attributed to the mayor to describe his approach: "in god we trust, all others bring data." There are, however, risks in this approach as well.

Consider the following map of Illinois's 4th congressional district in the Chicago area.



It is the result of *a data driven decision*. It would have been impossible to create this district without census data and an understanding of who would vote Democrat and/or Republican. Indeed, from the designer's perspective, the district is not only data driven, it is an *optimal outcome*. It achieves precisely the result it was designed to create, a predictable outcome for an election.

Data does not magically create outcomes that align with our values or our goals. Nor does it get used to foster utilitarian outcomes that automatically favor the greatest number of people. Power still matters. Environmental health and justice stakeholders have a responsibility to think about the behavior they model when using data and who they seek to empower and enable through their activities. In other worlds, values, goals and methodologies matter more and more in a world of data-sharing and require reflection and critical thinking.

Lesson #8: Sharing is about strategy, not an end in itself.

Finally, it is also important to recognize that those who do not agree with one's goals may also attempt to coopt or use one's data to advance their own goals. Better data does not remove the politics from an issue. It simply makes the data itself political.

The fact that there are risks around others using your data creates the ominous outcome of funders or grantees being paralyzed by fear and never sharing data. This should not be the case.

First, having opponents trying to discredit your data is an interesting barometer. It means they perceive you as a legitimate threat. In addition, any effort to discredit data can have two follow-on effects. First, it means that your opponents are entering into a discourse about data, one that could lead to a dialogue about what data IS actually accurate. This is a discourse that likely, in a longer term, is in a social movement's interests. Second is that engaging in your data will draw attention to your data and your perspective, raising your profile.

So even the risks can have upsides embedded in them. The key here is not to share data because it is the right thing to do, or to lock down data out of fear. It is to have a goal with your data: to know which organizations you want to mobilize or which outcomes you believe can be realized by sharing data. And, from the outset, to have thought about engagement strategies for how to deal with opponents and critics as much as with supporters

About this paper

This paper was authored by David Eaves as a consultant to HEFN, reflecting the views of the author. Eaves, a mediator and open innovation expert, is affiliated with the Berkman Center for Internet & Society at Harvard University.

This paper was produced for a June 2014 meeting on Advancing the Environmental Health Data Commons hosted by the Health and Environmental Funders Network (HEFN).

Comments may be addressed to the author at david@eaves.ca and to HEFN at hefn@hefn.org.

and “friendly” users.

Appendix: Examples of Environmental Health Data-Sharing

The following projects are examples of data-sharing efforts related to environmental health and justice. This list is intended to be illustrative, not comprehensive; in keeping with HEFN policy it is not intended as an endorsement of any project or grantseeker.

Chemical Hazard Data Commons is a collaboration among people who are interested in furthering the practices of chemical hazard assessment and alternatives assessment by building shared information resources. The project aims to freely provide collaboratively-curated data, to link together existing publicly available data using open technical standards, and to coordinate efforts to increase the knowledge base available to participants and the public.
To learn more: www.chemicalcommons.info

Health Product Declaration Collaborative is a customer-led organization committed to the continuous improvement of the building industry’s performance through transparency, openness and innovation in the product supply chain. The HPD Collaborative created and supports the Health Product Declaration (HPD), an impartial tool for the accurate reporting of product contents and each ingredient’s relationship to the bigger picture of human and ecological health. The HPD objectively defines the critical information needed to support accurate supply chain disclosure by manufacturers and suppliers, and informed decisions by building designers, specifiers, owners, and users.
To learn more: *http://hpdcollaborative.org*

Housing Checkup proposes to analyze public data on housing inspections for consumer-advocates, the real estate community, regulators, and policy-makers to improve the effectiveness, efficiency, and equity of housing quality regulations. For example, analyzing patterns of inspections and violations may identify underlying environmental factors that require neighborhood-wide interventions. Visualization of data can identify hotspots that require pro-active regulatory approaches. Comparing inspection statistics across cities would bolster benchmarks and performance standards that can be supported and re-enforced through policy. Analysis of inspection data could also help regulators deploy inspection resources more efficiently.

Housing Checkup leverages national efforts to make government housing inspection data more transparent. Recent developments in open data are bringing housing code inspections into the light. Several cities, such as Chicago, already maintain routinely updated public open data portals that include data on government service requests (e.g., 311 calls) and housing code violations. The House Facts Open Data Standard provides a uniform format for reporting government data on the operation, safety, and performance of residential buildings. Companies providing software for regulatory agencies, including Accela, are engineering their platforms to deliver open data. Other companies, such as Trulia and Civic Insight, are developing consumer-oriented applications to disclose property inspection data. Housing Checkup will complement these anticipated consumer applications of House Facts data with applications serving housing and consumer advocates, regulators, and policy-makers.
To learn more: *www.housingcheckup.org*

The FracTracker Alliance shares maps, data, and analyses to communicate impacts of the global oil and gas industry and inform actions that positively shape our energy future. Many people ask how many wells in the United States have been hydraulically fractured - or "fracked." It is an excellent question, but not one that is easily answered; most states don’t release data on well stimulation activities. Also, since the data are released by state regulatory agencies, it is necessary to obtain data from each state that has oil and gas data to even begin the conversation. While the endeavor is not complete, FracTracker recently found data for over 1.1 million active oil and gas wells in the U.S. and are in the process of determining which, if not all, are being fracked.
To learn more: *www.fractracker.org/2014/03/1-million-wells*

IVAN Online: Identifying Violations Affecting Neighborhoods. IVAN is an environmental monitoring system that connects the community with real people that can help solve local environmental problems. The IVAN Reporting Network currently includes sites for Coachella, Fresno, the Imperial Valley, Kern County, and Wilmington. Anyone may report an environmental issue of concern in the community. Reports may be viewed on a public map.
To learn more: *www.ivanonline.org*

The Public Laboratory for Open Technology and Science (Public Lab) is a community -- supported by a 501(c) 3 non-profit -- which develops and applies open-source tools to environmental exploration and investigation. By democratizing inexpensive and accessible Do-It-Yourself techniques, Public Lab creates a collaborative network of practitioners who actively re-imagine the human relationship with the environment.

The core Public Lab program is focused on "civic science" in which they research open source hardware and software tools and methods to generate knowledge and share data about community environmental health. Their goal is to increase the ability of underserved communities to identify, redress, remediate, and create awareness and

817 Silver Spring Avenue, Suite 110 • Silver Spring, MD 20910 • Tel: (301) 565-0500 • E-mail: info@hefn.org • www.HEFN.org

accountability around environmental concerns. Public Lab achieves this by providing online and offline training, education and support, and by focusing on locally-relevant outcomes that emphasize human capacity and understanding.
To learn more: www.publiclab.org

817 Silver Spring Avenue, Suite 110 • Silver Spring, MD 20910 • Tel: (301) 565-0500 • E-mail: info@hefn.org • www.HEFN.org