Recommendations for Equitable Open Data

Winter 2017

Detroit Digital Justice Coalition

Detroit Community Technology Project

Since the City of Detroit launched the Open Data Portal and GO DATA policy initiative in 2015, the Detroit Digital Justice Coalition (DDJC) and Detroit Community Technology Project (DCTP) have been exploring how to advance equitable practices for collecting, disseminating and using open data. By equitable practices, we mean accountable, ethical uses of public information for social good that actively resist the criminalization and surveillance of low income communities, people of color and other targeted communities.

In order to understand what equitable open data practices could look like in the City of Detroit, we asked a broad spectrum of Detroit residents to consider the potential benefits and harms of various data sets currently available on the City's Open Data Portal. From there, we assessed what actions can be taken by the City to maximize benefits and minimize harms, and investigated open data policies and tools in place in other cities that model our vision for data justice.
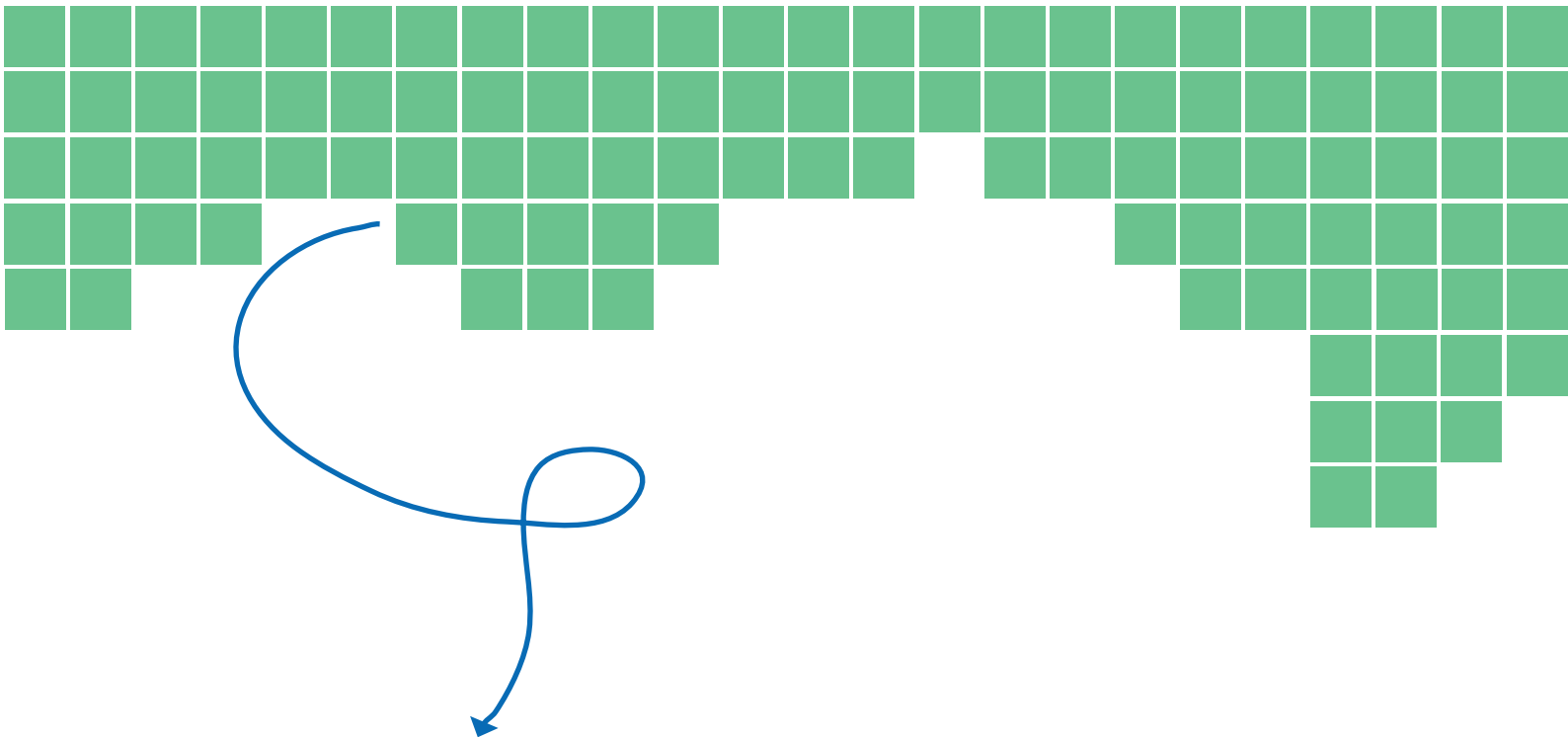
The recommendations we offer below are the outcome of our research and outline how the City of Detroit's Department of Innovation and Technology (DoIT) and other stewards of public information, in Detroit and beyond, can adopt and implement equitable open data practices.

We expect our recommendations to prompt critical conversations and questions, guide feature enhancements to Detroit's data portal and inform policy provisions to improve how institutions collect, disseminate and use open data. Along with our recommendations, we're also eager to share key takeaways from our research and resources for adapting our strategies.
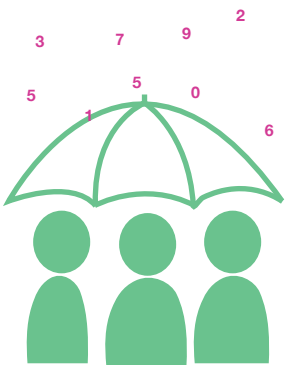
We anticipate that our recommendations, and the research that has shaped them, will continue to grow and evolve as we continue organizing around equitable open data.

This information is also presented as an open source project at
https://datajustice.github.io/report/

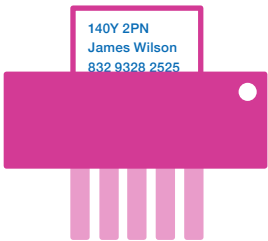We welcome feedback and your ideas at communitytech@alliedmedia.org.

Recommendations for Equitable Open Data in Detroit

Protect the people represented by the numbers

While open data positively boosts government transparency, it can also be a source of fear and harm to residents. While facilitating data justice scenarios, we heard deep-rooted concerns about how the Improve Detroit Open Issues queue can incite property swindles, like a metal scrapper pinpointing an abandoned vehicle, how Blight Tickets can reinforce "bad" and "undesirable" neighborhood reputations and influence property values or red-lining practices, or how a randomized block-level location of a Crime Incident isn't random enough in Detroit's most desolate areas to truly protect from re-victimization, stigmatization or further harassment.
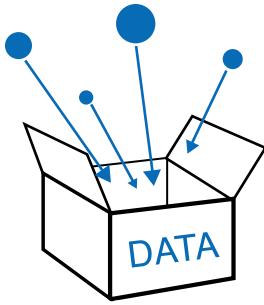
As the City continues to publish data, we ask for security audits to be performed and their outcomes shared by a third party, including penetration testing to expose potential vulnerabilities in existing data processing methods for anonymizing or randomizing personal data for release.

Don't retain personal information tied to accessing City services

We know that our personal information is often collected and stored when we apply for or access City services, such as entering a license plate number at a Park Detroit meter. While many privacy policies specifically promise to not share personal data with third parties, we want the City to go a step further and implement a provision to not retain any personal information associated with an application for City services. This would prevent the accidental loss or unauthorized use of personal data in ways that could harm or target residents.

New York City exemplifies how to roll out a successful Municipal ID program while pledging to not store cardholder's personal background information, a policy that both safeguards the personal information of their cardholders, who are more likely to be low income, experiencing homelessness, undocumented, or people of color, and illustrates their commitment to Sanctuary City status.

Publish data about all City services, even for privatized "public" services

The stated mission of the portal is, "to increase public access to valuable data and information concerning City government operations and service delivery." But, what about crucial City operations and services that are now administered by government contractors, private corporations, and quasi-public entities, like the Public Lighting Authority, Great Lakes Water Authority or Detroit Economic Growth Corporation? These entities are responsible for datasets that we found residents often cite as most important to have access to, such as utility shut-offs and development projects, yet are not subject to FoIA laws.

The portal should provide basic contract information about all service providers at minimum, and ideally additional data about their specific operating activities. For example, what vendor(s) does the City contract to remove garbage? Where is that vendor located, what is the timeline and total amount of their contract? The Sunlight Foundation offers legal language for a provision like this to ensure "government decisions to employ outside contractors do not result in the public losing access to its own information."

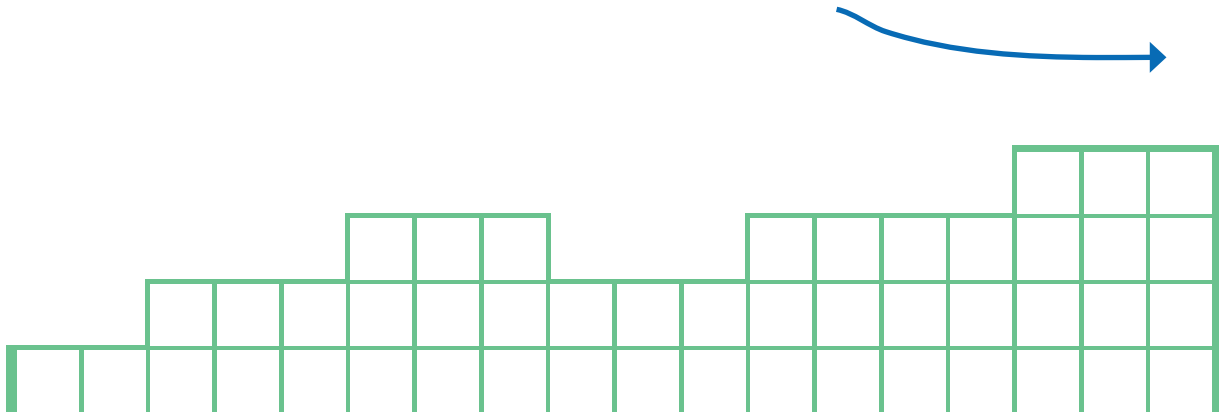### Prioritize the release of new datasets based on community interest

Currently, two key ways exist for residents to request public records not yet published to the portal: through the data portal's nomination form and through Freedom of Information Act (FoIA) requests. The first option requires internet access and appears to be a slow process, as only a single nomination from over 30 suggestions has been approved since launching. The second option, FoIAs, are also notoriously cumbersome and can be costly. Yet, these two mediums remain crucial ways for the City to gauge community interest and data needs.

FoIA requests and their responses, as well as approved nominated data sets, should be posted to the portal, to both boost relevant data sets for users and save internal resources in answering duplicate requests. We outlined our favorite ways that other cities are accomplishing this in our policy research.

Increase transparency around how data sets are defined and processed

While we appreciate all open data efforts, many datasets are not so easy to decipher. Participants in our data justice focus group often pointed to an example of confusing language in the Blight Violations dataset. How can we begin to use or analyze this dataset if we don't know how the City defines "blight"? Why is the definition not included with the description of the dataset? All data sets published on the Portal should include field definitions, collection methods, especially for anonymizing or randomizing personal data, and basic definitions of policy-specific or technical language.

The DPD: Citizen Complaints dataset exemplifies useful, thorough documentation by offering a description, contact information and field definitions, yet we've found many other data sets in Detroit's portal to be lacking these metadata. Beyond Detroit's portal, we also look to national cases where open data are being used for predictive analytics, like crime data being used for predictive sentencing in Philadelphia as expanded in our policy research. The algorithms and analysis methods associated with these open data are often inaccessible to the public and not subject to community oversight, yet can be encoded with implicit biases that affect how different factors are weighed. We ask that transparent practices apply to all open data definitions, methods, analysis and algorithms.
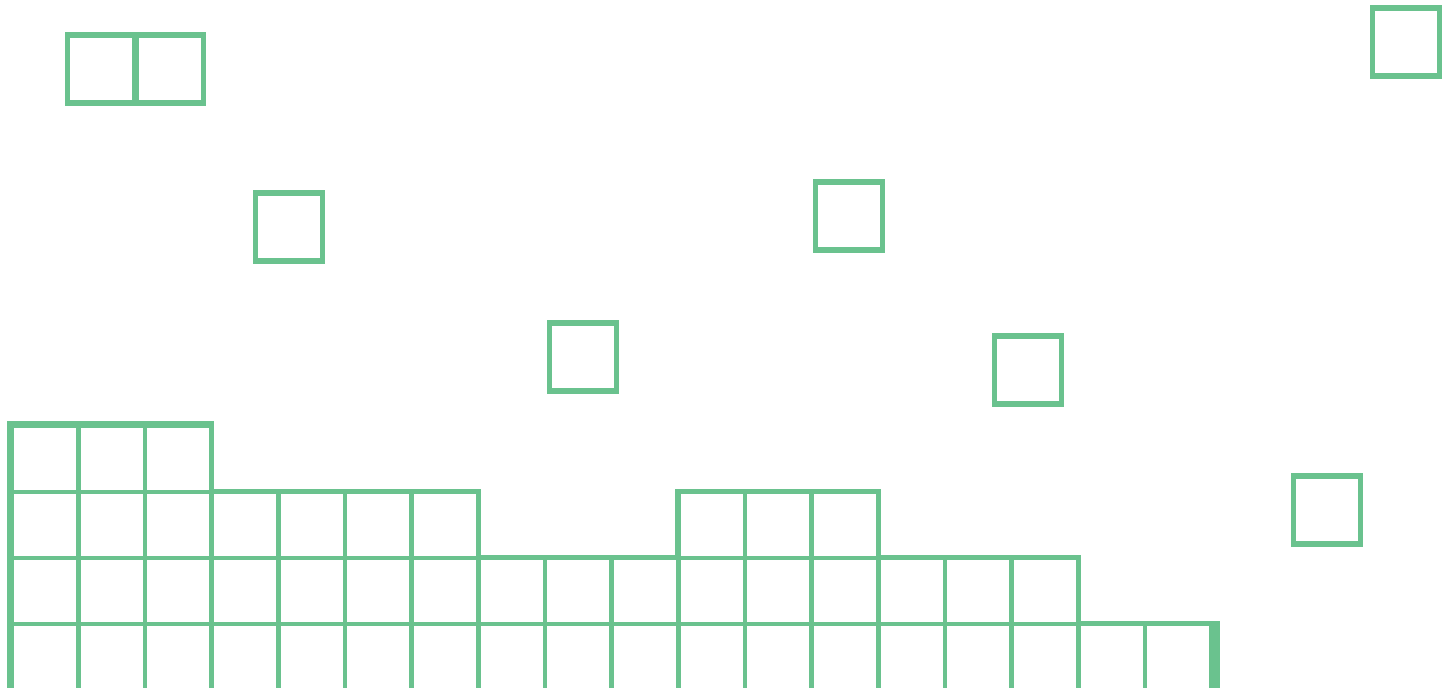
Engage residents offline about open data

In order to build awareness and open data literacy in Detroit, it's important to consider digital inequality and limited access to broadband internet. Through our data justice surveys, we found that most residents first learn about open data and the portal at events or workshops, but these can be infrequent and hard to access.

In addition to the online tutorials DoIT currently offers for learning to use the portal, we need an engagement plan that emphasizes community dialogue and in-person training sessions. We're sharing strategies we've found successful through organizing Data DiscoTechs, like partnering with local recreation centers and developing popular education methods to interactively learn about open data, as a starting point.

Share what's coming next

While it's simple to see how the Portal has grown since its launch through the Public Dataset Inventory, we also want to know what's coming next so that we can plan community research and organizing initiatives. Once or twice a year, DoIT should publish their key goals and activities, sharing with the public which dataset releases, portal feature improvements and public events to anticipate. Similar to the NYC Open Data Portal, the header of Detroit's portal should rotate to highlight features, identify the purpose, advertise new releases and offer quick links to things that matter most to Detroiters.

Research Strategies

Our recommendations for data justice presented above are the outcome of multiple participatory research projects and community education initiatives facilitated by the DDJC and DCTP around open data since early 2015.

We have collaboratively identified the following questions to guide and motivate our research:

✴ What should Detroit residents **be aware of** when it comes to open data?

✴ What are the **perceived benefits and harms** of open data to residents?

✴ How can we **inform all residents and educate our communities** about open data?

✴ What can we **learn from other cities**? What types of open data policies, portals and practices do they have in place?

✴ What does **participatory and democratic governance** of open data look like, and how can we help to facilitate it here in Detroit?

Below are summaries from our four key research projects, with links to follow to read more about our full methods and findings. Each research project has led us to understand the concept and practices of equitable open data more deeply and shaped our recommendations.

Surveys

The Data Justice Survey was created by members of the DDJC and is available to Data DiscoTech participants. The survey gauges familiarity with government data, perceptions of how open data is collected and used, and which data sets are most important for our communities to have access to. Read more.

Scenarios

Data Justice Scenarios are a collective brainstorming exercise facilitated through focus groups. Together, participants learn about a specific dataset, raise questions about how it is initially collected, and craft possible positive and negative use case scenarios of that data, with the goal of identifying actions and policies to maximize benefits and reduce harms. Read more.
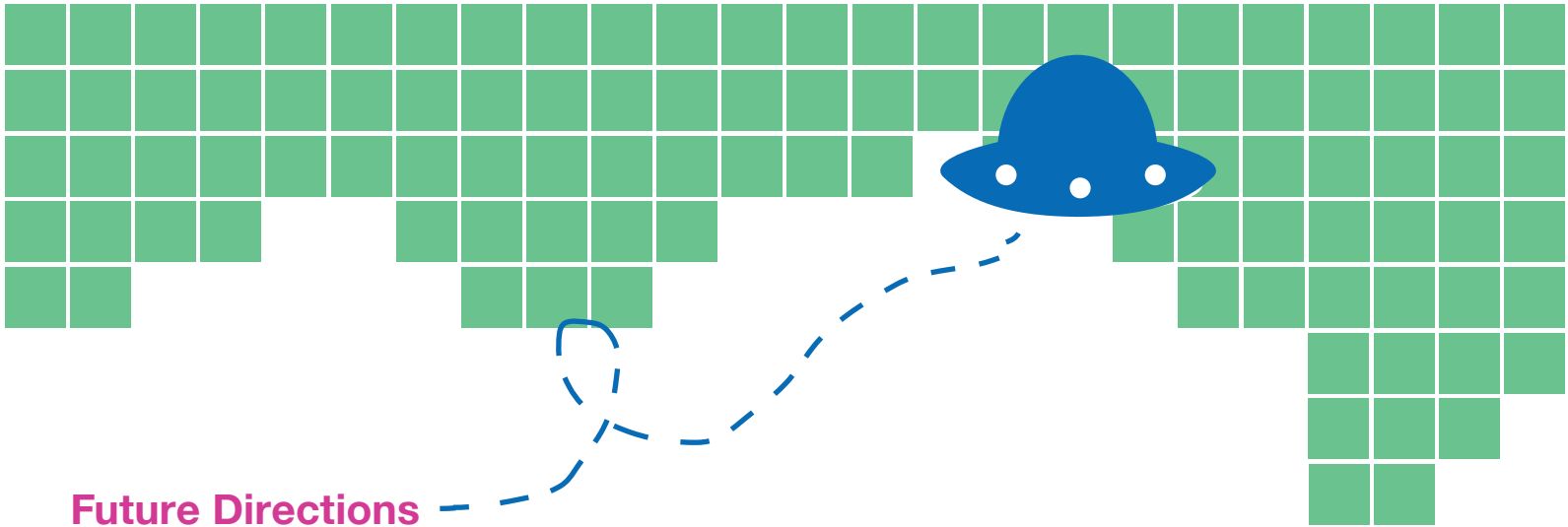
Policies, portals and practices in other cities

Research into open data policies has been completed by members of the DDJC and compares and contrasts what open data looks like and how it works across different U.S. cities, aiming to understand how data policy shapes data culture and highlight creative, just uses cases of open data. Read more.

Data DiscoTechs

Data DiscoTechs, or "discovering technology" fairs, are interactive workshops designed to demystify, engage and inform participants about the impact and possibilities of open data within our communities. Read more.

Future Directions

The DDJC and DCTP remain dedicated to exploring questions of transparency, accountability and data justice through participatory research and education initiatives, and aim to create paths for open data to be a source of community power. We'll continue to engage with our communities, the City and open data stakeholders to generate and advocate for the adoption of data justice policy provisions, and transform what we learn into reusable tools and curriculum to share with others seeding equitable open data initiatives in their cities.

For more on open data, check out these resources:

✴ Opening Data zine. Compiled by the DDJC and DCTP in 2015, the zine includes a primer on open data, real-world examples of data discrimination, use cases of data in organizing, creative data storytelling, and many links to follow for local projects.

✴ Data DiscoTech model. Through photos and videos, see what happens at a Data DiscoTech from the perspectives of participants and station managers.

✴ Open Data Policy Guidelines. Presented by the Sunlight Foundation, a national nonprofit dedicated to government transparency and accountability, these guidelines offer in-depth, technical ideas and language for comprehensive open data policy and are foundational to the locally-rooted recommendations we've presented above.