

**ONTARIO
SUPERIOR COURT OF JUSTICE
(DIVISIONAL COURT)**

BETWEEN:

CORPORATION OF THE CANADIAN CIVIL LIBERTIES ASSOCIATION
and LESTER BROWN

Applicants

and

TORONTO WATERFRONT REVITALIZATION CORPORATION, CITY OF
TORONTO, HER MAJESTY IN RIGHT OF ONTARIO as represented by the
MINISTER OF INFRASTRUCTURE, HER MAJESTY IN RIGHT OF
CANADA as represented by the MINISTER OF COMMUNITIES AND
INFRASTRUCTURE, AND THE ATTORNEY GENERAL OF CANADA

Respondents

APPLICATION under sections 2 and 6(1) and 6(2) of the *Judicial Review Procedure Act*, R.S.O. 1990, c. J.1, as amended, and sections 2, 7, 8 and 24 of the *Charter of Rights and Freedoms*.

AFFIDAVIT

I, Ben Green, of the City of Somerville, in the State of Massachusetts, in the United States of America, MAKE OATH AND SAY:

1. I am a PhD Candidate in Applied Math at the Harvard School of Engineering and Applied Sciences and an Affiliate at the Berkman Klein Center for Internet and Society at Harvard. I study the implementation and impacts of data science in local governments, with a focus on “smart cities” and the criminal justice system.

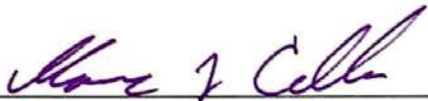
2. Attached here as **Exhibit "1"** is a copy of the report I have prepared in response to a request to give opinion evidence in this proceeding.

3. Attached to my report is the Acknowledgement of Expert's Duty that I have signed as well as my curriculum vitae outlining my education, experience and credentials.

4. The attached report accurately describes the instructions I received, the issues I was asked to address, my opinion respecting each issue and the reasons for my opinion. I have also described the factual assumptions on which my opinion is based, my research, and the documents I relied on in forming this opinion.

5. I believe that my report is accurate, based on the available information. I have prepared this report to the best of my ability.

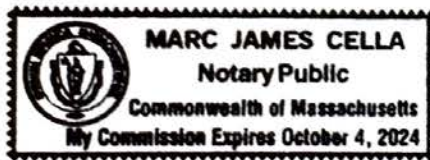
SWORN BEFORE ME at the City of
Cambridge in Massachusetts on May²⁴.,
2019



Commissioner for Taking Affidavits
(or as may be)



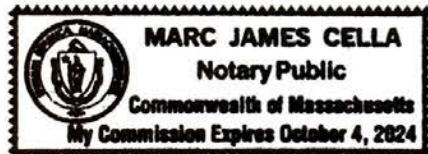
Ben Green



This is Exhibit "1" referred to in the Affidavit of Ben Green sworn
May 23rd 2019 MJC

Marc J Cella

Commissioner for Taking Affidavits (or as may be)



May 10, 2019

Young Park
Fogler, Rubinoff LLP
77 King Street West
Suite 3000, P.O. Box 95
TD Centre North Tower
Toronto, ON M5K 1G8

This report is structured into the following sections:

- A. Expert Qualifications
- B. Scope of Work, Instructions, and Assumptions
- C. Definitions
- D. Summary of Opinions
- E. Response to Question 1
- F. Response to Question 2
- G. Response to Question 3

Appendix A: References

Appendix B: Curriculum Vitae for Ben Green

A. Expert Qualifications

My name is Ben Green, and I am a scholar of municipal technology. I am a PhD Candidate at Harvard's School of Engineering and Applied Sciences and an Affiliate at the Berkman Klein Center for Internet and Society at Harvard.¹ My research focuses on the governance and social impacts of new technologies used by city governments. This research is informed by academic training as a data scientist, time spent working for the City of Boston as a data scientist, and collaborations with city data officers to develop effective and responsible privacy policies. Relevant publications include *The Smart Enough City*² (a book about opportunities and dangers of smart cities) and "Open Data Privacy"³ (a report for municipal officials about the privacy risks of data collection and use and about strategies for mitigating these dangers). I have a Master's degree in Applied Mathematics from Harvard University and a Bachelor's degree in Mathematics & Physics from Yale College.

My Curriculum Vitae is attached as Appendix B.

B. Scope of Work, Instructions, and Assumptions

You have asked me to provide opinions about Waterfront Toronto's and Sidewalk Labs' vision and plans with regard to Quayside. During the course of my review, I have reviewed documents

¹ I write only in my individual capacity, not on behalf of these organizations.

² Ben Green, *The Smart Enough City: Putting Technology in Its Place to Reclaim Our Urban Future* (MIT Press, 2019).

³ Ben Green et al., "Open Data Privacy: A risk-benefit, process-oriented approach to sharing and protecting municipal data," *Berkman Klein Center Research Publication* (2017), <http://nrs.harvard.edu/urn-3:HUL.InstRepos:30340010>

prepared by Waterfront Toronto and Sidewalk Labs pertaining to Quayside. I have assumed that these documents are accurate for the purpose of defining the Quayside project, and I will refer to these plans collectively as the “Quayside project” or “Quayside proposal.” Should these plans change, or should new evidence come to light, I reserve the right to revise or change my opinions stated here.

In order to complete this review, I have been asked to do the following:

1. To review documents pertaining to Waterfront Toronto’s and Sidewalk Labs’ plans.
2. On the basis of this information, to respond to three questions:
 - a. Are there grounds for concern about privacy breaches in respect of the captured private information, whether or not de-identified, and if there are grounds, do the materials reflect sufficient awareness of and preparation for such breaches?
 - b. Are the provisions for consent to the capture of private personal information within Quayside adequate, and will it be possible in light of the known facts to ensure meaningful consent by individuals to the capture and use of their personal private information?
 - c. Are the assurances that the captured private data will be secure adequate to allay concerns of its use in such a way as to violate privacy?

C. Definitions

Throughout this document, I will rely on the following definitions:

- De-identification: The process of altering a dataset to remove the identity of one or more individuals.
- Re-identification: The process of identifying the person or people represented in de-identified data.
- Personal information/Personally identifiable information (PII): Information about an identifiable individual. Examples include name, ID numbers, and medical records.
- Sensitive information: Data that includes intimate or identifiable information about one or more individuals. Personal information is a subset of sensitive information.
- Machine learning: Computer algorithms that learn from historical data to classify or make predictions about phenomena such as human behavior.

D. Summary of Opinions

Based on my review of the available information, I have developed the following opinions:

1. The Quayside project comes with many likely privacy harms, even among data that is labeled as “de-identified.” Based on my research, I believe that the current plans make promises about anonymity that are impossible to guarantee will be kept. Based on the current state of privacy technology and scholarship, I believe that Sidewalk Labs significantly understates the level of risk associated with the planned data collection.
2. I believe that the Quayside project lacks appropriate mechanisms for people to provide meaningful consent to the collection and use of information about them. Such consent is particularly difficult to provide in public spaces, where surveillance is so widespread as to be almost unavoidable simply by setting foot in Quayside.

3. The Quayside project includes the risk of many privacy harms that are possible given even perfectly secure data.

E. Question 1: Are there grounds for concern about privacy breaches in respect of the captured private information, whether or not de-identified, and if there are grounds, do the materials reflect sufficient awareness of and preparation for such breaches?

Based on the materials I have reviewed, the proposed privacy protections appear to rely on a definition of “personal information” that is unscientifically narrow and on assumptions about “de-identification” that overstate the capabilities of these methods. As a result, I believe that the Quayside project overstates the extent to which it is able to rely on non-sensitive “de-identified” data.

It has been well known for many years that traditional distinctions between sensitive and non-sensitive data fail to capture the underlying reality: supposedly anonymous or de-identified data can reveal intimate details about individuals and populations.⁴

The most common distinction is based on the presence of “personal information” or “personally identifiable information” (PII)—features such as name and ID number that, on their own, identify individuals. Data containing PII is considered to be sensitive and hence worthy of protection, while data without PII is not.⁵ While PII is sensitive, a great deal of data that does not contain PII is also sensitive.⁶

This distinction breaks down in two primary ways, as supposedly anonymous or de-identified data can be combined with other information or analyzed in search of patterns to reveal sensitive information.⁷

Breakdown 1: Seemingly innocuous data can be combined with other information to reveal sensitive information.

This first issue is often known as the “mosaic effect,” for the way in which different datasets can be combined to form a mosaic that identifies individuals and reveals private information. A notable example of the mosaic effect occurred in 1997, when Massachusetts Governor William Weld released state employee medical records for research purposes, promising that the information was anonymous. A few days later, however, Weld received a letter: it contained his own health records, culled from the released data.⁸ The envelope came from Latanya Sweeney, then a graduate student at the Massachusetts Institute of Technology, who had identified Weld’s file by linking the medical records with publicly-available voter lists.⁹ Although the medical

⁴ Paul M. Schwartz and Daniel J. Solove, “The PII Problem: Privacy and a New Concept of Personally Identifiable Information,” *NYU Law Review* 86 (2011).

⁵ Schwartz and Solove, “The PII Problem.”

⁶ Schwartz and Solove, “The PII Problem.”

⁷ Green et al., “Open Data Privacy”.

⁸ Erica Klarreich, “Privacy by the Numbers: A New Approach to Safeguarding Data,” *Quanta Magazine* (2012), <https://www.quantamagazine.org/a-mathematical-approach-to-safeguarding-private-data-20121210/>

⁹ Latanya Sweeney, “Simple Demographics Often Identify People Uniquely,” (Carnegie Mellon University, Data Privacy Working Paper 3, 2000).

records did not directly reveal anyone's identity, it contained each patient's zip code, birthday, and sex—information that is also listed in publicly available and identified voter lists—allowing Sweeney to match the records and identify which medical records belonged to Weld.

This is just one of many examples of the mosaic effect leading sensitive information to being revealed from data that on its face seems de-identified. Another notable example came in 2008 when two computer scientists re-identified “anonymized” data released by Netflix, combining that data with public information available on the Internet Movie Database (IMDB) to identify the people behind certain records and infer their political preferences and other potentially sensitive information.¹⁰ In 2006, AOL released “anonymized” search histories, which journalists were able to re-identify by searching the internet for related and publicly available information.¹¹

As more and more data is collected by different actors, it becomes increasingly difficult to promise that any dataset—even one that appears to be anonymized—will not be re-identifiable when combined with other information.¹² For example, in 2014, in response to a Freedom of Information Law (FOIL) request, New York City released data detailing every taxi ride recorded in registered NYC taxis during 2013.¹³ The data was meant to be anonymized and contained information about pickup time and location, drop-off time and location, and the taxicab (in the form of license plate) and driver (in the form of medallion number) involved in each trip. Yet by combining that data with published paparazzi photos, a data scientist found that it was possible to identify where celebrities photographed getting into cabs were going.¹⁴ By combining this taxi data with any published report of someone's location (e.g., paparazzi photos, a Facebook or Instagram post), it can be possible to track where specific individuals are traveling.

Breakdown 2: Data analysis of seemingly innocuous data can find or infer patterns that reveal sensitive information.

The second issue occurs when detailed records about people's movements or other behaviors—precisely the type of data collected by smart city infrastructure (e.g., sensors and cameras)—are analyzed to reveal sensitive information. One single data point—for instance, a phone's location at a particular place at a particular time—is unlikely to reveal someone's identity or anything sensitive about them. But when millions of data points are collected and combined with modern analysis techniques—which Sidewalk Labs has stated an intention to do throughout its vision statement (see e.g., page 72)¹⁵—such data can be used to track people's behavior and infer a great deal about them. Consider the following examples.

¹⁰ Arvind Narayanan and Vitaly Shmatikov, "Robust De-anonymization of Large Sparse Datasets" (paper presented at the IEEE Symposium on Security and Privacy, 2008).

¹¹ Michael Barbaro and Tom Zeller Jr., "A Face Is Exposed for AOL Searcher No. 4417749," *The New York Times* (2006), <https://www.nytimes.com/2006/08/09/technology/09aol.html>

¹² Green et al., "Open Data Privacy".

¹³ Chris Whong, "FOILING NYC's Taxi Trip Data," (2014), http://chriswhong.com/open-data/foil_nyc_taxi/ (Accessed May 14, 2019).

¹⁴ Anthony Tockar, "Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset," *Neustar Research* (2014), <https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/> (Accessed May 14, 2019).

¹⁵ Sidewalk Labs, "Vision Sections of RFP Submission," (2017), <https://sidewalktoronto.ca/wp-content/uploads/2017/10/Sidewalk-Labs-Vision-Sections-of-RFP-Submission.pdf> (Accessed May 14, 2019).

Granular data about people's activities is so sensitive in aggregate—despite each record seeming so benign in isolation—because human behavior is “highly unique.”¹⁶ Data collected on a massive scale captures each person's unique patterns of behavior. Two studies demonstrated this phenomenon by analyzing datasets that contained the mobile phone location traces¹⁷ and credit card transactions¹⁸ of more than one million people. Even though both datasets lacked PII—they included just a random number corresponding to each person as well as the locations and times that those people were tracked—it was possible to identify individuals and learn about their behavior. Remarkably, more than 90 percent of people could be uniquely identified with just four data points of where they have been and when they were there.¹⁹ In other words, just because someone may be in a crowd while going about their day, their specific set of locations visited is likely to be unique—and hence identifiable.

Another example demonstrates how aggregate data about urban patterns—especially data tied to locations in the city—can be analyzed to reveal sensitive information about individuals. As described above, New York City released data detailing every taxi ride recorded in registered NYC taxis during 2013.²⁰ The data contained information about the pickup and drop-off times and locations of each trip. By analyzing the destinations of all the trips leaving from a specific location, it was possible to identify the home addresses of several patrons of a Manhattan strip club.²¹ Then via the mosaic effect, it was possible to combine this information with other information that is available publicly online to identify the names of these patrons. The same method of looking for patterns in all the taxi trips originating or ending at a specific location could be used to learn who attends almost any location in the city. Indeed, similar data about trips on London's bike share program was analyzed using related methods to reveal the travel patterns of several individuals, including where those individuals live and work.²²

Furthermore, when data is combined with machine learning it is even possible to infer a great deal of personal information that is not explicitly contained in a dataset. With detailed information about where you have been, for instance, machine learning algorithms can predict whom you know²³ and where you will go next.²⁴ Algorithms can detect whether someone is depressed based on the photos she posts on Instagram.²⁵ Data about seemingly routine behaviors

¹⁶ Yves-Alexandre de Montjoye et al., "Unique in the Crowd: The privacy bounds of human mobility," *Nature* *503* (2013).

¹⁷ de Montjoye et al., "Unique in the Crowd."

¹⁸ Yves-Alexandre de Montjoye et al., "Unique in the shopping mall: On the reidentifiability of credit card metadata," *Science* *347*, no. 6221 (2015).

¹⁹ de Montjoye et al., "Unique in the Crowd."; de Montjoye et al., "Unique in the shopping mall."

²⁰ Whong, "FOILING NYC's Taxi Trip Data".

²¹ Tockar, "Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset".

²² James Siddle, "I Know Where You Were Last Summer: London's public bike data is telling everyone where you've been," *The Variable Tree* (2014), <https://vartree.blogspot.co.uk/2014/04/i-know-where-you-were-last-summer.html> (Accessed May 14, 2019).

²³ Nathan Eagle, Alex Sandy Pentland, and David Lazer, "Inferring friendship network structure by using mobile phone data," *Proceedings of the National Academy of Sciences* *106*, no. 36 (2009).

²⁴ Lars Backstrom, Eric Sun, and Cameron Marlow, "Find Me If You Can: Improving Geographical Prediction with Social and Spatial Proximity" (paper presented at the Proceedings of the 19th International Conference on World Wide Web, 2010).

²⁵ Andrew G. Reece and Christopher M. Danforth, "Instagram photos reveal predictive markers of depression," *EPJ Data Science* *6*, no. 1 (2017).

such as Facebook Likes can reveal sexual identity, race, political affiliation, and even whether one's parents are married.²⁶ These results are not uncommon, and instead reflect the broader trend that large datasets create the ability to “produce novel insights that probably couldn't have been revealed in any other way.”²⁷

As a result of these two breakdowns just described—that data can be re-identified through the mosaic effect and through various data analysis techniques—many researchers, public bodies, and others have recognized that existing distinctions about “anonymous” or “personal” data are inadequate. Below are just three examples:

- The President's Council of Advisors on Science and Technology (PCAST): “By data mining and other kinds of analytics, non-obvious and sometimes private information can be derived from data that, at the time of their collection, seemed to raise no, or only manageable, privacy issues. [...] one can never know what information may later be extracted from any particular collection of big data.”²⁸
- Computer scientists Arvind Narayanan and Vitaly Shmatikov: “The versatility and power of re-identification algorithms imply that terms such as ‘personally identifiable’ and ‘quasi-identifier’ simply have no technical meaning. While some attributes may be uniquely identifying on their own, any attribute can be identifying in combination with others.”²⁹
- Legal scholar Paul Ohm: “Data can either be useful or perfectly anonymous but never both. [...] reidentification science exposes the underlying promise made by [privacy] laws—that anonymization protects privacy—as an empty one.”³⁰

Sidewalk Labs

In its various planning and presentation materials, Sidewalk Labs promises to make much of its data “non-personal” and “de-identified,” describing this as data that is “designed not to trace back to any individual.”³¹ But based on the research just described, I do not believe that this is a promise that can be reliably kept. As a result, I believe that Sidewalk Labs is drastically understating the level of privacy harms involved in their proposed data collection practices.

A look at Sidewalk Labs' LinkNYC program provides an instructive example of how the company's proposed privacy plans in Quayside fail to address the known privacy harms of de-identified data. LinkNYC is a partnership between New York City and Sidewalk Labs to provide

²⁶ Michal Kosinski, David Stillwell, and Thore Graepel, “Private traits and attributes are predictable from digital records of human behavior,” *Proceedings of the National Academy of Sciences* 110, no. 15 (2013).

²⁷ Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt, 2013).

²⁸ President's Council of Advisors on Science and Technology, “Big Data and Privacy: A Technological Perspective,” (2014).

²⁹ Arvind Narayanan and Vitaly Shmatikov, “Myths and fallacies of personally identifiable information,” *Communications of the ACM* 53, no. 6 (2010).

³⁰ Paul Ohm, “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization” *UCLA Law Review* 57 (2009).

³¹ Sidewalk Labs, “Digital Governance Proposals for DSAP Consultation,” (2018), https://waterfrontoronto.ca/nbe/wcm/connect/waterfront/41979265-8044-442a-9351-e28ef6c76d70/18.10.15_SWT_Draft+Proposals+Regarding+Data+Use+and+Governance.pdf?MOD=AJPERES (Accessed May 10, 2019).

free public Wi-Fi via more than 7,500 internet-connected kiosks placed throughout NYC.³² In its proposals for Toronto, Sidewalk Labs has emphasized “ubiquitous connectivity” and suggested that it plans to integrate similar kiosks into Quayside.³³

Per its privacy policy, LinkNYC kiosks are equipped with sensors and cameras that gather data about everyone in its vicinity.³⁴ And while LinkNYC promises strong privacy protections for “Personally Identifiable Information,” that information is defined primarily as a user’s name and email address. Other information, such as a device’s MAC address (a device’s unique identifier that helps it connect to the internet), operating system, and other device identifiers, are collected and considered “Technical Information.” And while the kiosks do not collect anyone’s precise location, they do know everyone’s general location based on the location of each kiosk that is collecting information. Such location data is often highly re-identifiable and can reveal particularly sensitive information about individuals.³⁵ Moreover, research has found that even coarse location data (such as a kiosk’s location rather than a person’s precise location in that vicinity) is only marginally less sensitive than precise location data.³⁶

The research and examples discussed above explain how this “technical information” is in fact quite sensitive and can reveal intimate information about individuals. Yet LinkNYC treats this data as non-personal, asserting that it may share the information with advertisers and others. In turn, the reliance on a narrow definition of personal data means that a great deal of potentially sensitive information is considered to be innocuous and is not properly safeguarded.

Many privacy advocates and experts have critiqued LinkNYC’s privacy policies for these reasons.³⁷ One privacy lawyer has stated that the LinkNYC privacy policy is designed “to make you believe that something is being promised, when actually it lets them do anything they want.”³⁸ Similarly, New York Civil Liberties Union’s executive director argued, “Free public Wi-Fi can be an invaluable resource for this city, but New Yorkers need to know there are too many strings attached.”³⁹

In sum, I believe that there is significant reason to be concerned about privacy harms associated with the collected data, even data that is defined as “de-identified.” Any data comes with risks of re-identification. I do not believe that Sidewalk Labs has adequately addressed the risk

³² City of New York Office of the Mayor, “Mayor de Blasio Announces Public Launch of LinkNYC Program, Largest and Fastest Free Municipal Wi-Fi Network in the World,” (2016), <http://www1.nyc.gov/office-of-the-mayor/news/184-16/mayor-de-blasio-public-launch-linknyc-program-largest-fastest-free-municipal#/0> (Accessed May 14, 2019).

³³ Sidewalk Labs, “Vision Sections of RFP Submission”.

³⁴ LinkNYC, “Privacy Policy,” (2017), <https://www.link.nyc/privacy-policy.html> (Accessed May 14, 2019).

³⁵ Green et al., “Open Data Privacy”.

³⁶ de Montjoye et al., “Unique in the Crowd.”

³⁷ See for example Nick Pinto, “Google Is Transforming NYC’s Payphones Into a ‘Personalized Propaganda Engine’,” *The Village Voice* (2016), <https://www.villagevoice.com/2016/07/06/google-is-transforming-nycs-payphones-into-a-personalized-propaganda-engine/> (Accessed May 14, 2019); Benjamin Dean, “The Heavy Price We Pay for ‘Free’ Wi-Fi,” *Government Technology* (2016), <https://www.govtech.com/opinion/The-Heavy-Price-We-Pay-for-Free-Wi-Fi.html> (Accessed May 14, 2019).

³⁸ Pinto, “Google Is Transforming NYC’s Payphones”.

³⁹ New York Civil Liberties Union, “City’s Public Wi-Fi Raises Privacy Concerns,” (2016), <https://www.nyclu.org/en/press-releases/nyclu-citys-public-wi-fi-raises-privacy-concerns> (Accessed May 14, 2019).

associated with de-identified data, instead making promises about anonymity that are impossible to guarantee will be kept. This significantly understates the level of risk associated with their data collection plans.

F. Question 2: Are the provisions for consent to the capture of private personal information within Quayside adequate, and will it be possible in light of the known facts to ensure meaningful consent by individuals to the capture and use of their personal private information?

In my opinion, the Quayside project lacks appropriate mechanisms for those whose data is being collected and used to provide meaningful consent.

The Office of the Privacy Commissioner of Canada has written guidelines about what meaningful consent entails.⁴⁰ The guidelines indicate that:

- Information provided about the collection, use and disclosure of individuals' personal information must be readily available in complete form.
- Information must be provided to individuals in manageable and easily-accessible ways.
- Individuals cannot be required to consent to the collection, use or disclosure of personal information beyond what is necessary to provide the product or service – they must be given a choice.

The General Data Protection Regulation (GDPR) contains similar requirements regarding consent.⁴¹ The GDPR requires that consent must be:

- Freely given: Consent must be given on a voluntary basis involving real choice.
- Informed: People should be notified who is collecting data, what data will be collected, and how that data will be stored and used. People must also be informed about the right to withdraw consent at any time (and the process of withdrawing consent must be as easy as providing consent).
- Specific: Consent must be bound to specified purposes.
- Unambiguous: Consent requires either a statement or a clear affirmative act. Consent cannot be implied and must always be given through an opt-in.

It is useful to analyze the Quayside project according to these four requirements from the GDPR, which encompass the requirements stated by the Canadian Office of the Privacy Commissioner.

Freely Given

In my opinion, consent cannot be freely given in Quayside because the project involves collecting data about everyone in a public space, regardless of whether those people have consented to that data being collected. In its own materials and presentations, Sidewalk Labs has acknowledged that “meaningful consent cannot be reasonably or reliably achieved” in public

⁴⁰ Office of the Privacy Commissioner of Canada, "Guidelines for obtaining meaningful consent," (2018), https://www.priv.gc.ca/en/privacy-topics/collecting-personal-information/consent/gl_omc_201805 (Accessed May 14, 2019).

⁴¹ Official Journal of the European Union, "General Data Protection Regulation," (2016), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679> (Accessed May 14, 2019).

spaces.⁴² Many people will be sharing their data in Quayside because they have no choice but to share that data in order to spend time in Quayside, not necessarily because they genuinely want to share that information.

Once smart city technologies such as sensors and cameras are installed (Sidewalk Labs articulates an intention to use these and other technologies⁴³), I believe that it will be almost impossible for someone to avoid being tracked while in Quayside. Sensors will monitor the behavior of anyone with a Bluetooth or Wi-Fi connected device. And given the expansive reach of cameras (and the growing use and availability of facial recognition software), it is increasingly impossible to escape being tracked even by abandoning one's personal digital technology. Based on the documents I have reviewed, the pervasive tracking also means that information about minors will be collected in Quayside alongside the information about everyone else.

This reality leaves the public with a stark choice: accept widespread surveillance or stay out of Quayside altogether. This is an impossible choice, especially for those who live or work in Quayside. Traveling through Quayside will require "accepting" Sidewalk Labs' terms simply by setting foot in the neighborhood. I believe that it is impossible for consent to be freely given if people must visit Quayside to engage fully in civic society and are unable to avoid having their behavior tracked while there.

Informed

It is impossible for the public to give informed consent when they lack complete information about what data is being collected and how it will be used and shared. The privacy policies that people sign before using online services are designed to be complex and do not provide a detailed description of risk.⁴⁴ It would take people 76 days per year (reading eight hours per day) to read every privacy policy they encounter in a given year.⁴⁵

In Quayside, Sidewalk Labs has proposed placing hexagonal icons in public spaces where data is being collected.⁴⁶ This is a useful step toward informing the public what type of information is being collected in a given location.

Yet I believe that the proposed hexagons are still not fully informative. In particular, this proposal lacks two areas of information that are essential to informed consent:

1. Risks: As described in response to Question 1, research indicates that there is not a binary classification of identifiable and de-identified data—yet the hexagons divide data into these two camps. As a result, a hexagon showing that de-identified data is being collected understates the potential risks of that data being collected. While the data itself may be

⁴² Sidewalk Labs, "Digital Governance Proposals for DSAP Consultation".

⁴³ Sidewalk Labs, "Vision Sections of RFP Submission".

⁴⁴ Woodrow Hartzog, "User Agreements Are Betraying You," *Medium: Trust Issues* (2018), <https://medium.com/s/trustissues/user-agreements-are-betraying-you-19db7135441f> (Accessed May 8, 2019).

⁴⁵ Alexis C. Madrigal, "Reading the Privacy Policies You Encounter in a Year Would Take 76 Work Days," *The Atlantic* (2012), <https://www.theatlantic.com/technology/archive/2012/03/reading-the-privacy-policies-you-encounter-in-a-year-would-take-76-work-days/253851/>

⁴⁶ Jacqueline Lu, "How can we bring transparency to urban tech? These icons are a first step.," *Medium: Sidewalk Talk* (2019), <https://medium.com/sidewalk-talk/how-can-we-make-urban-tech-transparent-these-icons-are-a-first-step-f03f237f8f00> (Accessed May 10, 2019).

labeled as “de-identified,” that implies to the public that the data is truly anonymous when in fact research has shown that “any attribute can be identifying in combination with others.”⁴⁷

2. Storage and use: Privacy harms occur not simply when data is collected, but also in the various ways that that data is stored and used. Some of this information is available through the digital interface following the QR code on the hexagons, but the proposed project fails to fully describe all of the ways that one’s data may be shared and analyzed.

There also appear to be no mechanisms for withdrawing consent, such as by having Sidewalk Labs stop collecting data or remove all the data it has collected about an individual. Once Sidewalk Labs collects data about someone, it appears that that person no longer has any ability to control how that information is used.

Specific

Based on my research, consent in the Quayside proposal is very broad, concerning all-or-nothing data collection practices. For example, there appear to be no mechanisms to accept certain types of data collection or certain uses of that data but to deny others.

Unambiguous

I believe that consent in Quayside is quite ambiguous, as people will have data about them collected whenever they enter Quayside. Sidewalk Labs’ processes do not appear to include provisions for people to provide any unambiguous affirmative consent before they are subject to data collection.

In sum, I believe that the Quayside project lacks appropriate mechanisms for those whose data is being collected and used to provide meaningful consent. In particular, the Quayside proposal fails to provide for consent that is freely given, informed, specific, and unambiguous (as per the Canadian guidelines and the GDPR). Sidewalk Labs proposes providing some information to the public about the data being collected—but even if these practices were sufficient to inform the public, I do not believe that a fully-informed public is necessarily a consenting public. Consent must also be freely given, specific, and unambiguous. Thus, knowing how one’s information is being collected and used does not mean that one consents to those practices. This is especially so in public spaces, where people may have no choice but to accept broad data collection if they want or need to spend time in those public spaces.

G. Question 3: Are the assurances that the captured private data will be secure adequate to allay concerns of its use in such a way as to violate privacy?

Insecure data is not the only potential harm related to data collection. Even perfectly secure data (which is impossible⁴⁸) can still raise several significant privacy risks.

⁴⁷ Narayanan and Shmatikov, “Myths and fallacies of personally identifiable information.”

⁴⁸ Ira S. Rubinstein and Woodrow Hartzog, “Anonymization and Risk,” *Washington Law Review* 91 (2016).

An essential aspect of security is ensuring that only those who are supposed to have access to data are in fact able to access that information.⁴⁹ Yet I believe the most salient privacy risks of the Quayside project are related to authorized—rather than unauthorized—data access and use.

Based on my review of the documents, I believe that there are many harms to the public of Sidewalk Labs and its partners collecting and retaining data gathered about individuals. While each individual data point may not on its own be particularly sensitive, the aggregation of millions of data points provides these companies with a great deal of knowledge and power over the public.

The most familiar concern regarding privacy is widespread surveillance, allowing governments and companies to watch your every action and expose secrets. Such fears tap into deep-seated cultural notions about privacy that are drawn from Big Brother, the totalitarian government from George Orwell's novel *1984*. Following Orwell's influence, writes leading privacy scholar Daniel Solove in *The Digital Person*, we typically conceive of privacy following the "secrecy paradigm": the idea that privacy is invaded when one's secrets are observed, leading people to self-censor (via "chilling effects") or suffer the consequences.⁵⁰

The concept of Big Brother captures some elements of why privacy is essential for maintaining civil liberties. For example, Toronto police monitored Black Lives Matter activists during demonstrations in 2016.⁵¹ Such surveillance can have significant chilling effects on free speech and assembly: several studies have found that surveillance causes people to limit their speech and the information they view online.⁵²

Yet the concept of Big Brother cannot fully explain why privacy is important.⁵³ As we have already seen, a great deal of data collection today relies on information that is neither secret, illegal, nor embarrassing—in fact, each individual data point appears meaningless and anonymous. The secrecy paradigm thus fails to explain the harms of someone's bike share trips or Facebook Likes being collected, aggregated, and analyzed. For as Solove explains, nowadays many uses of data "aim not at suppressing individuality but at studying it and exploiting it."⁵⁴

People today have little knowledge or control regarding what personal data is collected, who owns it, and how they use it.⁵⁵ As more data is collected and used by governments and companies, privacy becomes defined less by the secrets that any piece of information reveals and increasingly by the inferences that large amounts of relatively non-sensitive data make

⁴⁹ National Research Council, *Computers at Risk: Safe Computing in the Information Age* (The National Academies Press, 1991).

⁵⁰ Daniel J Solove, *The Digital Person: Technology and Privacy in the Information Age* (NYU Press, 2004).

⁵¹ Stephen Davis, "Police monitored Black Lives Matter Toronto protesters in 2016, documents show," *CBC* (2018), <https://www.cbc.ca/news/canada/toronto/police-monitored-black-lives-matter-toronto-protesters-in-2016-documents-show-1.4645628> (Accessed May 14, 2019).

⁵² Karen Gullo, "Surveillance Chills Speech—As New Studies Show—And Free Association Suffers," *Electronic Frontier Foundation* (2016), <https://www.eff.org/deeplinks/2016/05/when-surveillance-chills-speech-new-studies-show-our-rights-free-association> (Accessed May 12, 2019).

⁵³ Solove, *The Digital Person*.

⁵⁴ Solove, *The Digital Person*.

⁵⁵ Solove, *The Digital Person*; Bruce Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World* (W. W. Norton & Company, 2015).

possible—and the power that those inferences grant. For example, Facebook can calibrate its News Feed algorithm to influence a user's mood⁵⁶ and likelihood to vote.⁵⁷ The dating site OkCupid can alter profile match scores to affect certain people's chances of getting dates.⁵⁸ Healthcare services can deny coverage if they learn that someone recently visited websites associated with having cancer.⁵⁹

Although data collection touches everyone, the most severe impacts of diminishing privacy are suffered by racialized groups and the poor. Despite being more concerned about privacy than their more well-off counterparts, most lower-income individuals lack the knowledge of privacy settings and policies to sufficiently reduce how much they are tracked.⁶⁰ And given that activists in racial justice movements like Black Lives Matters are targeted for surveillance⁶¹ and undocumented immigrants face deportation,⁶² racialized groups are prone to suffer the greatest consequences of being identified and tracked by the government.

Racialized groups and the poor are also most susceptible to harms due to a lack of privacy in dealings with private companies. As companies increasingly make decisions about individuals using data drawn from their online behavior and social networks, lower socioeconomic groups can be unfairly excluded from credit, jobs, housing, and healthcare.⁶³ Low-wage workplaces monitor their employees' keystrokes, location, emails, and online browsing to detect unsanctioned behavior, which can result in firing.⁶⁴

Thus, as security expert Bruce Schneier explains in his book *Data and Goliath*, many companies already possess the knowledge and influence necessary to restrict individual autonomy and exploit people.⁶⁵ Smart cities technologies will vastly increase the scale and scope of data that tech companies collect. Companies that place cameras and sensors on Wi-Fi kiosks, trashcans, and streetlights will gain heretofore impossible—and highly profitable—insights about the behavior of individuals.

⁵⁶ Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences* 111, no. 24 (2014).

⁵⁷ Robert M. Bond et al., "A 61-million-person experiment in social influence and political mobilization," *Nature* 489, no. 7415 (2012).

⁵⁸ Christian Bautista, "OKCupid: 'We experiment on human beings,'" *Tech Times* (2014), <https://www.techtimes.com/articles/11475/20140730/okcupid-we-experiment-on-human-beings.htm> (Accessed May 14, 2019).

⁵⁹ Casey Johnston, "Denied for that loan? Soon you may thank online data collection," *Ars Technica* (2013), <https://arstechnica.com/business/2013/10/denied-for-that-loan-soon-you-may-thank-online-data-collection/> (Accessed May 14, 2019).

⁶⁰ Mary Madden et al., "Privacy, Poverty and Big Data: A Matrix of Vulnerabilities for Poor Americans," *Washington University Law Review* 95 (2017).

⁶¹ Davis, "Police monitored Black Lives Matter Toronto protesters in 2016, documents show".

⁶² Kathleen Harris, "Canada Border Services Agency moves to 'substantially' increase deportations," *Canadian Broadcasting Corporation* (2018), <https://www.cbc.ca/news/politics/cbsa-deportations-border-removals-1.4873169> (Accessed May 14, 2019).

⁶³ John Podesta et al., *Big Data: Seizing Opportunities, Preserving Values* (Executive Office of the President, 2014).

⁶⁴ "The rise of workplace spying," *The Week* (2015), <http://theweek.com/articles/564263/rise-workplace-spying> (Accessed May 14, 2019).

⁶⁵ Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*.

Those with limited financial resources are more vulnerable to the tracking of smart city companies. For example, while well-off New Yorkers who do not want LinkNYC to track them can forgo free Wi-Fi in favor of a personal data plan, lower-class residents have no alternative to free Wi-Fi (indeed, the whole point of LinkNYC is to provide internet access to those who cannot afford it). They must accept being tracked in exchange for Internet access, essentially giving up their privacy in order to access basic services and infrastructure.

These privacy harms can be exacerbated as data collected by one company (in this case, Sidewalk Labs) is shared with the government and other companies.⁶⁶

One way that information can be shared widely across companies is through data brokers, companies that aggregate data about people and sell or share that information with other companies.⁶⁷ There are more than 4,000 data broker companies worldwide, and overall the data brokerage industry is a \$200 billion industry.⁶⁸ Given the vast reach of data brokers that gather and share data without the public's knowledge or consent, one company's data can easily end up in the hands of another.⁶⁹ The profiles created by these data brokers (such as "suffering seniors" and "urban scramble") make it possible for companies to target predatory loans and scams with precision.⁷⁰ These harms can arise as companies either share their own data with data brokers or acquire data from data brokers to combine it with their own data.

Another way that sensitive information can be revealed is through "open data" initiatives, which involve releasing datasets online in an effort to make government more transparent and accountable and to foster collaboration with the public.⁷¹ Because data collected in cities relates to the people within those cities, open data can—through the mosaic effect and data analysis—reveal sensitive information about individuals. By releasing open data, cities have inadvertently revealed the identities of sexual assault victims⁷² and people who carry large sums of cash at night,⁷³ as well as people's medical information,⁷⁴ political affiliation,⁷⁵ and travel patterns and where they live and work.⁷⁶

⁶⁶ Green et al., "Open Data Privacy".

⁶⁷ Federal Trade Commission, "Data Brokers: A Call for Transparency and Accountability," (Washington DC: Federal Trade Commission, 2014).

⁶⁸ WebFX, "What Are Data Brokers – And What Is Your Data Worth?," <https://www.webfx.com/blog/general/what-are-data-brokers-and-what-is-your-data-worth-infographic/>

⁶⁹ Federal Trade Commission, "Data Brokers: A Call for Transparency and Accountability."

⁷⁰ Federal Trade Commission, "Data Brokers: A Call for Transparency and Accountability."

⁷¹ Green et al., "Open Data Privacy".

⁷² Andrea Peterson, "Why the names of six people who complained of sexual assault were published online by Dallas police," *The Washington Post* (2016), <https://www.washingtonpost.com/news/the-switch/wp/2016/04/29/why-the-names-of-six-people-who-complained-of-sexual-assault-were-published-online-by-dallas-police/> (Accessed May 14, 2019); Gabe Cunningham, "Crime Open Data & Reverse Geocoding," (2016), <https://medium.com/@gabecunn/crime-open-data-reverse-geocoding-493f88e725ba> (Accessed May 14, 2019).

⁷³ Claudia Vargas, "City settles gun permit posting suit," *The Philadelphia Inquirer* (2014), http://www.philly.com/philly/news/local/20140723_City_settles_gun_permit_suit_for_1_4_million.html

⁷⁴ Klarreich, "Privacy by the Numbers: A New Approach to Safeguarding Data".

⁷⁵ Ethan Chiel, "Why the D.C. government just publicly posted every D.C. voter's address online," *Splinter* (2016), <https://splinternews.com/why-the-d-c-government-just-publicly-posted-every-d-c-1793857534> (Accessed May 14, 2019).

⁷⁶ Siddle, "I Know Where You Were Last Summer: London's public bike data is telling everyone where you've been".

Although there are strategies that cities can take to reduce the risks of such disclosures, there is nonetheless an inevitable tension between open data's utility (more detailed data provides greater transparency and can be used for more purposes) and risks (more detailed data contains more sensitive information)—a tension that will grow only more severe as the scope of municipal data collection expands.⁷⁷

Although open data programs efforts are typically operated by city governments, Sidewalk Labs has itself expressed an intention to create its own “open data hub” as part of the Quayside project.⁷⁸ They write, “By default, non-personal Urban Data will be open and freely accessible to the public.”⁷⁹ This practice could lead to significant privacy harms. Given that it is impossible to know what data will be sensitive when analyzed or combined with other data (see Section 1), it is impossible for Sidewalk Labs to ensure that the data it releases as open data will not be used in ways that violate privacy and that could harm the public. As open data is made public, stalkers, abusers, law enforcement, and companies could all employ this data to take advantage of others. Access and use of Sidewalk's “de-identified” data is likely to be further streamlined by its proposed Application Programming Interface (API) that will streamline access to data for developers.⁸⁰

It is also common for data collected by smart city companies to be shared with local law enforcement.⁸¹ For example, LinkNYC's privacy policy notes that it may share data and video footage may be shared with government agencies as required by law (such as in response to a subpoena or court order) or to investigate any incidents related to the kiosks. Law enforcement officials across the United States made over 130,000 requests for access to digital evidence from just six technology companies in 2017, and more than 80% of these requests were granted by the companies.⁸² If similar policies are adopted (or if existing law requires sharing data with law enforcement), any data collected by Sidewalk Labs in Quayside could end up in the hands of law enforcement—and thus that Quayside will not simply be a home of corporate data collection, but also potentially a home of widespread police surveillance. This can lead to many harms, such as the oppression of racialized groups and political dissidents.

Any information shared with the government (in Canada⁸³ as well as in the US) is also prone to being released through freedom of information (FOI) and public records laws, which compel the government to release data they control when requested to do so by a member of the public. Although these laws in both countries contain exemptions restricting the release of personal information, at least in the United States these laws' reliance on the outdated PII framework

⁷⁷ Green et al., “Open Data Privacy”.

⁷⁸ Sidewalk Labs, “Digital Governance Proposals for DSAP Consultation”.

⁷⁹ Sidewalk Labs, “Digital Governance Proposals for DSAP Consultation”.

⁸⁰ Sidewalk Labs, “Vision Sections of RFP Submission”.

⁸¹ William A. Carter and Jennifer C. Daskal, “Low-Hanging Fruit: Evidence-Based Solutions to the Digital Evidence Challenge,” (2018), https://csis-prod.s3.amazonaws.com/s3fs-public/publication/180725_Carter_DigitalEvidence.pdf?tAGR_DvxRdp0RspiGYNGcGKTUjrGY3rN (Accessed May 8, 2019).

⁸² Carter and Daskal, “Low-Hanging Fruit: Evidence-Based Solutions to the Digital Evidence Challenge”.

⁸³ I base this statement on my interpretation of the Ontario Municipal Freedom of Information and Protection of Privacy Act and the Ontario Freedom of Information and Protection of Privacy Act.

severely limits what data a government can withhold for being too sensitive. While information including someone's health history is likely to be protected against public release, for example, datasets containing records that are not directly identifiable but pertain to the behavior of individuals are unlikely to be seen as sufficiently personal to be exempt from release as public records. For example, the NYC taxi trip data that was analyzed to infer the identities of strip club patrons was initially released in 2014 through a public records request and then posted online by the requestor for anyone to use.⁸⁴ Similarly, in 2013, the ACLU used a public records request to obtain three years of license plate readings from the Seattle Police Department (containing 7.3 million scans recording a license plate, location, and time), data that provided sufficient information to track the daily travel patterns of both police officers and the public.⁸⁵ And as cities gather and store more supposedly-anonymous data about people's behavior, they will be sitting on increasingly large piles of information that are prone to public release but could expose sensitive information about individuals.

In sum, my opinion is that the Quayside project involves many privacy harms that are possible even given perfectly secure data. All of the data sharing and harms just described result from legal and standard data sharing practices. As security expert Bruce Schneier puts it, "data is a toxic asset and saving it is dangerous."⁸⁶ Once data is collected, it is prone to be released. New technology that allows more granular and sensitive information to be collected magnifies these risks.

⁸⁴ Green et al., "Open Data Privacy".

⁸⁵ Jamela Debelak, "ALPR: The Surveillance Tool You've Probably Never Heard Of," *ACLU of Washington* (2013), <https://aclu-wa.org/blog/alpr-surveillance-tool-you-ve-probably-never-heard> (Accessed May 14, 2019).

⁸⁶ Bruce Schneier, "Data is a toxic asset, so why not throw it out?," *CNN* (2016), <http://www.cnn.com/2016/03/01/opinions/data-is-a-toxic-asset-opinion-schneier/index.html> (Accessed May 9, 2019).

Appendix A

Primary Documents

- Lu, Jacqueline. "How can we bring transparency to urban tech? These icons are a first step." *Medium: Sidewalk Talk* (2019). <https://medium.com/sidewalk-talk/how-can-we-make-urban-tech-transparent-these-icons-are-a-first-step-f03f237f8ff0> (Accessed May 10, 2019).
- Sidewalk Labs. "Digital Governance Proposals for DSAP Consultation." (2018). https://waterfrontoronto.ca/nbc/wcm/connect/waterfront/41979265-8044-442a-9351-e28ef6c76d70/18.10.15_SWT_Draft+Proposals+Regarding+Data+Use+and+Governance.pdf?MOD=AJPERES (Accessed May 10, 2019).
- Sidewalk Labs. "Vision Sections of RFP Submission." (2017). <https://sidewalktoronto.ca/wp-content/uploads/2017/10/Sidewalk-Labs-Vision-Sections-of-RFP-Submission.pdf> (Accessed May 14, 2019).

Other References

- Backstrom, Lars, Eric Sun, and Cameron Marlow. "Find Me If You Can: Improving Geographical Prediction with Social and Spatial Proximity." Paper presented at the Proceedings of the 19th International Conference on World Wide Web, 2010.
- Barbaro, Michael, and Tom Zeller Jr. "A Face Is Exposed for AOL Searcher No. 4417749." *The New York Times* (2006). <https://www.nytimes.com/2006/08/09/technology/09aol.html>
- Bautista, Christian. "OKCupid: 'We experiment on human beings'." *Tech Times* (2014). <https://www.techtimes.com/articles/11475/20140730/okcupid-we-experiment-on-human-beings.htm> (Accessed May 14, 2019).
- Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. "A 61-million-person experiment in social influence and political mobilization." *Nature* 489, no. 7415 (2012): 295-98.
- Carter, William A., and Jennifer C. Daskal. "Low-Hanging Fruit: Evidence-Based Solutions to the Digital Evidence Challenge." (2018). https://csis-prod.s3.amazonaws.com/s3fs-public/publication/180725_Carter_DigitalEvidence.pdf?tAGR_DvxRdp0RspiGYNGcGK_TUjrGY3rN (Accessed May 8, 2019).
- Chiel, Ethan. "Why the D.C. government just publicly posted every D.C. voter's address online." *Splinter* (2016). <https://splinternews.com/why-the-d-c-government-just-publicly-posted-every-d-c-1793857534> (Accessed May 14, 2019).
- City of New York Office of the Mayor. "Mayor de Blasio Announces Public Launch of LinkNYC Program, Largest and Fastest Free Municipal Wi-Fi Network in the World." (2016). <http://www1.nyc.gov/office-of-the-mayor/news/184-16/mayor-de-blasio-public-launch-linknyc-program-largest-fastest-free-municipal#/0> (Accessed May 14, 2019).
- Cunningham, Gabe. "Crime Open Data & Reverse Geocoding." (2016). <https://medium.com/@gabecunn/crime-open-data-reverse-geocoding-493f88e725ba> (Accessed May 14, 2019).
- Davis, Stephen. "Police monitored Black Lives Matter Toronto protesters in 2016, documents show." *CBC* (2018). <https://www.cbc.ca/news/canada/toronto/police-monitored-black->

- [lives-matter-toronto-protesters-in-2016-documents-show-1.4645628](#) (Accessed May 14, 2019).
- de Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. "Unique in the Crowd: The privacy bounds of human mobility." *Nature* 3 (2013).
- de Montjoye, Yves-Alexandre, Laura Radaelli, Vivek Kumar Singh, and Alex "Sandy" Pentland. "Unique in the shopping mall: On the reidentifiability of credit card metadata." *Science* 347, no. 6221 (2015): 536-39.
- Dean, Benjamin. "The Heavy Price We Pay for 'Free' Wi-Fi." *Government Technology* (2016). <https://www.govtech.com/opinion/The-Heavy-Price-We-Pay-for-Free-Wi-Fi.html> (Accessed May 14, 2019).
- Debelak, Jamela. "ALPR: The Surveillance Tool You've Probably Never Heard Of." *ACLU of Washington* (2013). <https://aclu-wa.org/blog/alpr-surveillance-tool-you-ve-probably-never-heard> (Accessed May 14, 2019).
- Eagle, Nathan, Alex Sandy Pentland, and David Lazer. "Inferring friendship network structure by using mobile phone data." *Proceedings of the National Academy of Sciences* 106, no. 36 (2009): 15274-78.
- Federal Trade Commission. "Data Brokers: A Call for Transparency and Accountability." Washington DC: Federal Trade Commission, 2014.
- Green, Ben. *The Smart Enough City: Putting Technology in Its Place to Reclaim Our Urban Future*. MIT Press, 2019.
- Green, Ben, Gabe Cunningham, Ariel Ekblaw, Paul Kominers, Andrew Linzer, and Susan Crawford. "Open Data Privacy: A risk-benefit, process-oriented approach to sharing and protecting municipal data." *Berkman Klein Center Research Publication* (2017). <http://nrs.harvard.edu/urn-3:HUL.InstRepos:30340010>
- Gullo, Karen. "Surveillance Chills Speech—As New Studies Show—And Free Association Suffers." *Electronic Frontier Foundation* (2016). <https://www.eff.org/deeplinks/2016/05/when-surveillance-chills-speech-new-studies-show-our-rights-free-association> (Accessed May 12, 2019).
- Harris, Kathleen. "Canada Border Services Agency moves to 'substantially' increase deportations." *Canadian Broadcasting Corporation* (2018). <https://www.cbc.ca/news/politics/cbsa-deportations-border-removals-1.4873169> (Accessed May 14, 2019).
- Hartzog, Woodrow. "User Agreements Are Betraying You." *Medium: Trust Issues* (2018). <https://medium.com/s/trustissues/user-agreements-are-betraying-you-19db7135441f> (Accessed May 8, 2019).
- Johnston, Casey. "Denied for that loan? Soon you may thank online data collection." *ArsTechnica* (2013). <https://arstechnica.com/business/2013/10/denied-for-that-loan-soon-you-may-thank-online-data-collection/> (Accessed May 14, 2019).
- Klarreich, Erica. "Privacy by the Numbers: A New Approach to Safeguarding Data." *Quanta Magazine* (2012). <https://www.quantamagazine.org/a-mathematical-approach-to-safeguarding-private-data-20121210/>
- Kosinski, Michal, David Stillwell, and Thore Graepel. "Private traits and attributes are predictable from digital records of human behavior." *Proceedings of the National Academy of Sciences* 110, no. 15 (2013): 5802-05.

- Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. "Experimental evidence of massive-scale emotional contagion through social networks." *Proceedings of the National Academy of Sciences* 111, no. 24 (June 17, 2014): 8788-90.
- LinkNYC. "Privacy Policy." (2017). <https://www.link.nyc/privacy-policy.html> (Accessed May 14, 2019).
- Madden, Mary, Michele E. Gilman, Karen Levy, and Alice E. Marwick. "Privacy, Poverty and Big Data: A Matrix of Vulnerabilities for Poor Americans." *Washington University Law Review* 95 (2017): 53–125.
- Madrigal, Alexis C. "Reading the Privacy Policies You Encounter in a Year Would Take 76 Work Days." *The Atlantic* (2012).
<https://www.theatlantic.com/technology/archive/2012/03/reading-the-privacy-policies-you-encounter-in-a-year-would-take-76-work-days/253851/>
- Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt, 2013.
- Narayanan, Arvind, and Vitaly Shmatikov. "Myths and fallacies of personally identifiable information." *Communications of the ACM* 53, no. 6 (2010): 24-26.
- Narayanan, Arvind, and Vitaly Shmatikov. "Robust De-anonymization of Large Sparse Datasets." Paper presented at the IEEE Symposium on Security and Privacy, 2008.
- National Research Council. *Computers at Risk: Safe Computing in the Information Age*. The National Academies Press, 1991.
- New York Civil Liberties Union. "City's Public Wi-Fi Raises Privacy Concerns." (2016).
<https://www.nyclu.org/en/press-releases/nyclu-citys-public-wi-fi-raises-privacy-concerns> (Accessed May 14, 2019).
- Office of the Privacy Commissioner of Canada. "Guidelines for obtaining meaningful consent." (2018). https://www.priv.gc.ca/en/privacy-topics/collecting-personal-information/consent/gl_omc_201805 (Accessed May 14, 2019).
- Official Journal of the European Union. "General Data Protection Regulation." (2016).
<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679> (Accessed May 14, 2019).
- Ohm, Paul. "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization". *UCLA Law Review* 57 (2009): 1701.
- Peterson, Andrea. "Why the names of six people who complained of sexual assault were published online by Dallas police." *The Washington Post* (2016).
<https://www.washingtonpost.com/news/the-switch/wp/2016/04/29/why-the-names-of-six-people-who-complained-of-sexual-assault-were-published-online-by-dallas-police/> (Accessed May 14, 2019).
- Pinto, Nick. "Google Is Transforming NYC's Payphones Into a 'Personalized Propaganda Engine'." *The Village Voice* (2016). <https://www.villagevoice.com/2016/07/06/google-is-transforming-nycs-payphones-into-a-personalized-propaganda-engine/> (Accessed May 14, 2019).
- Podesta, John, Penny Pritzker, Ernest J. Moniz, John Holdren, and Jeffrey Zients. *Big Data: Seizing Opportunities, Preserving Values*. Executive Office of the President, 2014.
- President's Council of Advisors on Science and Technology. "Big Data and Privacy: A Technological Perspective." 2014.
- Reece, Andrew G., and Christopher M. Danforth. "Instagram photos reveal predictive markers of depression." *EPJ Data Science* 6, no. 1 (2017).

- Rubinstein, Ira S., and Woodrow Hartzog. "Anonymization and Risk." *Washington Law Review* 91 (2016): 703.
- Schneier, Bruce. *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. W. W. Norton & Company, 2015.
- Schneier, Bruce. "Data is a toxic asset, so why not throw it out?" *CNN* (2016). <http://www.cnn.com/2016/03/01/opinions/data-is-a-toxic-asset-opinion-schneier/index.html> (Accessed May 9, 2019).
- Schwartz, Paul M., and Daniel J. Solove. "The PII Problem: Privacy and a New Concept of Personally Identifiable Information." *NYU Law Review* 86 (2011): 1814.
- Siddle, James. "I Know Where You Were Last Summer: London's public bike data is telling everyone where you've been." *The Variable Tree* (2014). <https://vartree.blogspot.co.uk/2014/04/i-know-where-you-were-last-summer.html> (Accessed May 14, 2019).
- Solove, Daniel J. *The Digital Person: Technology and Privacy in the Information Age*. NYU Press, 2004.
- Sweeney, Latanya. "Simple Demographics Often Identify People Uniquely." Carnegie Mellon University, Data Privacy Working Paper 3, 2000.
- The Week Staff. "The rise of workplace spying." *The Week* (2015). <http://theweek.com/articles/564263/rise-workplace-spying> (Accessed May 14, 2019).
- Tockar, Anthony. "Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset." *Neustar Research* (2014). <https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/> (Accessed May 14, 2019).
- Vargas, Claudia. "City settles gun permit posting suit." *The Philadelphia Inquirer* (2014). http://www.philly.com/philly/news/local/20140723_City_settles_gun_permit_suit_for_1_4_million.html
- WebFX. "What Are Data Brokers – And What Is Your Data Worth? ." <https://www.webfx.com/blog/general/what-are-data-brokers-and-what-is-your-data-worth-infographic/>
- Whong, Chris. "FOILING NYC's Taxi Trip Data." (2014). http://chriswhong.com/open-data/foil_nyc_taxi/ (Accessed May 14, 2019).

Ben Green

Harvard University
Maxwell Dworkin 209
33 Oxford St, Cambridge, MA 02138

Phone: (617) 413-0594
Email: bgreen@g.harvard.edu
Site: <http://scholar.harvard.edu/bgreen>

- INTERESTS** Data, algorithms, and social justice
Municipal governance of technology
- AFFILIATIONS** **Berkman Klein Center for Internet & Society at Harvard**
Affiliate 2018 – Present
Fellow 2016 – 2018
- EDUCATION** **Harvard University**
PhD in Applied Mathematics 2020 (expected)
MS in Applied Mathematics 2016
- Yale University**
BS in Mathematics & Physics, with distinction (Cum Laude) 2014
- GRANTS** Berkman Klein Center for Internet & Society Fellowship 2016
Harvard Kennedy School Taubman Center Urban Experience Fellowship 2016
NSF Graduate Research Fellowship 2015
DOD National Defense Science and Engineering Graduate Fellowship (declined) 2015
Herbert Winokur SEAS Graduate Fellowship 2015
Eric & Wendy Schmidt Data Science for Social Good Summer Fellowship 2014
Dwight Hall at Yale Urban Fellowship 2013
New Haven Mayor’s Community Arts Grant 2013
Yale President’s Public Service Fellowship 2013
Alan S. Tetelman 1958 Fellowship for International Research in the Sciences 2011
- BOOKS** Ben Green. *The Smart Enough City: Putting Technology in Its Place to Reclaim Our Urban Future*. MIT Press. (2019).
- PAPERS** Ben Green and Yiling Chen. “Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments.” *ACM Conference on Fairness, Accountability, and Transparency (ACM FAT*)* (2019). **Best Technical and Interdisciplinary Paper**
- Ben Green. “‘Fair’ Risk Assessments: A Precarious Approach for Criminal Justice Reform.” *5th Workshop on Fairness, Accountability, and Transparency in Machine Learning (ICML)* (2018).
- Ben Green and Lily Hu. “The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning.” *Machine Learning: The Debates Workshop (ICML)* (2018).
- Ben Green, Thibaut Horel, and Andrew Papachristos. “Modeling contagion through social networks to explain and predict gunshot violence in Chicago, 2006 to 2014.” *JAMA Internal Medicine* 177, no. 3 (2017): 326–333.
- Ben Green, Gabe Cunningham, Ariel Ekblaw, Paul Kominers, Andrew Linzer, and Susan Crawford. “Open Data Privacy: A risk-benefit, process-oriented approach to sharing and protecting municipal data,” *Berkman Klein Center Research Publication* (2017).
- Ben Green, Paul Bardunias, J. Scott Turner, Radhika Nagpal, and Justin Werfel. “Excavation and aggregation as organizing factors in de novo construction by mound-building termites.” *Proceedings of the Royal Society B* 284, no. 1856 (2017).

Ben Green, Alejandra Caro, Matt Conway, Robert Manduca, Tom Plagge, and Abby Miller. "Mining administrative data to spur urban revitalization." *Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)* (2015).

Ben Green. "Testing and quantifying collective intelligence," *Proceedings of the Collective Intelligence Conference* (2015).

SELECTED TALKS

2019

MIT Press Bookstore
MIT Department of Urban Studies and Planning
Harvard Institute for Learning in Retirement
Carleton University Master of Public Policy and Administration program
Harvard Applied Computation 221: Critical Thinking in Data Science (guest lecture)
Seton Hall University Law School Institute for Privacy Protection Spring Conference
University of California, Irvine
Harvard Sociology 98AB: Urban Politics in the Global City (guest lecture)
Berkman Klein Center Luncheon Series
ACM FAT*
Crime Lab New York
AI Now Institute, NYU

2018

Strategic Leadership Development for Senior Vietnamese Government Officials
MetroLab Annual Summit (panel moderator)
University of Indiana Ostrom Workshop on Smart Cities
Privacy Task Force for New Jersey Municipalities
FATML (ICML workshop)
Machine Learning: The Debates (ICML workshop)
Humboldt University of Berlin Faculty of Law
Boston City Council (invited expert testimony)
Berkman Klein Center Attorney General Tech Forum
Seton Hall Law School Artificial Intelligence and the Law Conference

2017

Seton Hall Law School Institute for Privacy Protection Conference on New and Nontraditional Actors in Privacy and Social Media Regulation
Cambridge City Council (invited expert testimony)
National Network for Safe Communities National Conference
Harvard Data Privacy Lab Talks on Technology Science
LibrePlanet
Boston Area Research Initiative Spring Conference
Future of Privacy Forum Smart Cities working group

2016

City of Cambridge Open Data Review Board
Digital Communities Mid-Year CIO Leadership Group Meeting

2015

KDD
Collective Intelligence

RESEARCH EXPERIENCE

Harvard University
Computer Science Department Graduate research assistant
Criminal justice algorithms September 2017 – Present
Studying the social impacts of risk assessments in the criminal justice system.

Berkman Center for Internet & Society Data governance fellow
Best practices for municipal data governance January 2016 – August 2017
Developed best practices for how cities manage data and technology. Studied the privacy

implications behind open data and developing a framework for assessing privacy risks when sharing data. Provided resources for cities to protect against discrimination when making data-driven decisions. Regularly convened with and presented to municipal leaders.

Yale University

Sociology Department

Gun violence in co-offending networks

Research assistant

January 2014 – January 2017

Studied the structure of criminal networks in eight American cities and identified risk factors for gunshot victims. Analyzed police records on arrests and shootings to model the diffusion of gun violence as an epidemic that spreads from person to person via social interactions. Developed a predictive model for who is at risk to be shot that outperforms traditional approaches.

Harvard University

Computer Science Department

Collective intelligence in termite colonies

Graduate research assistant

September 2014 – May 2016

Studying collective intelligence in termite colonies to determine how termites self-organize to collectively construct mounds. Designed experiments and conducted field research in Namibia. Developed simulations to infer the social dynamics in self-organizing groups of termites.

The Eric & Wendy Schmidt

Data Science for Social Good

Summer Fellowship

Data mining for urban revitalization

Research fellow

June 2014 – August 2014

Worked with the Mayor's Innovation Team in Memphis, TN to identify data-driven strategies for urban revitalization. Developed a machine learning classifier and interactive website to help policymakers and developers identify distressed houses in Memphis.

Yale University

Physics Department

Improved sampling of galaxy clustering

Undergraduate senior thesis

September 2013 – May 2014

Analyzed and developed algorithms and statistical methods to produce accurate sampling of galaxy clusters for the Dark Energy Spectroscopic Instrument.

Yale University

Mechanical Engineering Department

Emergent group behavior of insect swarms

Research assistant

September 2013 – January 2014

Studied the emergent behavior and complex dynamics of insect swarms. Used network applications to analyze the interactions between pairs of insects.

CERN

Statistical tests to detect elementary particles

Research assistant

May 2011 – July 2011

Worked on the ATLAS experiment of the Large Hadron Collider. Analyzed decay patterns of top quarks to search for a Z boson outside of the Standard Model. Conducted statistical analyses of particle collisions, comparing Monte Carlo simulations with recorded ATLAS data.

PROFESSIONAL EXPERIENCE

City of Boston

Department of Innovation & Technology

Municipal data analytics and policy

Data analytics fellow

June 2016 – May 2017

Worked for the Citywide Analytics Team analyzing data and developing policies to aid City Departments improve operations and services. Analyzed Fire Department and EMS responses and made recommendations for process improvements, including a pilot program that pairs public health and medical resources to respond to certain incidents. Aided in the development of policies and practices for a new open data portal.

City of New Haven

Department of Transportation

Improving transportation efficiency and safety

Policy intern

May 2013 – May 2014

Analyzed New Haven's on-street parking regulations and made changes in order to reduce con-

gestion and aid economic development. Coordinated adoption of cellphone payment technology in meters throughout the city. Conceived and initiated process of creating a traffic garden for New Haven. Wrote pedestrian and bicycle safety guides.

Design for America at Yale

Creating artistic bike racks

Created a team to promote a more sustainable cycling environment in New Haven. Initiated and ran a program matching local artists and businesses to create three downtown bike racks that double as public art. Received a 2013 New Haven Mayor's Community Arts Grant to fund artistic bike racks throughout New Haven.

Team founder and leader

September 2012 – May 2014

Litl, Inc.

Machine learning for computer vision

Developed machine learning and computer vision algorithms for the photo-viewing application Woven. Developed a classifier to determine whether a picture was taken indoors or outdoors. Used techniques such as logistic regression, graph clustering, and Bayesian analysis.

Research and development intern

May 2012 – August 2012

TEACHING

Faculty member, UC Irvine Technology, Law, and Society Summer Institute, June 2018.

Course assistant, Harvard Law School Responsive Communities Lab, Fall 2016.

Head teaching fellow, Harvard Computer Science 182: Artificial Intelligence, Fall 2015.

Math and science coordinator, Dwight Hall Academic Mentoring Program at Yale.

Tutor, Yale College Science and Quantitative Reasoning Center.

SERVICE

Program Committee: Black in AI (NeurIPS workshop) 2018; Conference on Fairness, Accountability, and Transparency (FAT*) 2019, International ACM Web Science Conference (WebSci) 2019, Debugging Machine Learning Models (ICLR workshop) 2019, Mechanism Design for Social Good (EC workshop) 2019

Reviews: MIT Press (3x); Big Data & Society; Online Information Review; Data Mining and Knowledge Discovery; npj Digital Medicine

Institutional: Harvard Graduate Student Union Bargaining Committee Member

**ONTARIO
SUPERIOR COURT OF JUSTICE
(DIVISIONAL COURT)**

BETWEEN:

CORPORATION OF THE CANADIAN CIVIL LIBERTIES ASSOCIATION and
LESTER BROWN

Applicants

and

TORONTO WATERFRONT REVITALIZATION CORPORATION, CITY OF
TORONTO, HER MAJESTY IN RIGHT OF ONTARIO as represented by the
MINISTER OF INFRASTRUCTURE, HER MAJESTY IN RIGHT OF CANADA as
represented by the MINISTER OF COMMUNITIES AND INFRASTRUCTURE, AND
THE ATTORNEY GENERAL OF CANADA

Respondents

APPLICATION under sections 2 and 6(1) and 6(2) of the *Judicial Review Procedure Act*, R.S.O. 1990,
c. J.1, as amended, and sections 2, 7, 8 and 24 of the *Charter of Rights and Freedoms*.

FORM 53

ACKNOWLEDGMENT OF EXPERT'S DUTY

1. My name is Ben Green. I live in Sommerville in the state of Massachusetts.
2. I have been engaged by or on behalf of the Corporation of the Canadian Civil Liberties Association and Lester Brown to provide evidence in relation to the above-noted court proceeding.
3. I acknowledge that it is my duty to provide evidence in relation to this proceeding as follows:
 - (a) to provide opinion evidence that is fair, objective and non-partisan;
 - (b) to provide opinion evidence that is related only to matters that are within my area of expertise; and
 - (c) to provide such additional assistance as the court may reasonably require, to determine a matter in issue.
4. I acknowledge that the duty referred to above prevails over any obligation which I may owe to any party by whom or on whose behalf I am engaged.

Date ...May 24, 2019.....



Signature

CORPORATION OF THE CANADIAN CIVIL LIBERTIES
ASSOCIATION et al.
Applicants

-and- TORONTO WATERFRONT REVITALIZATION
CORPORATION et al.
Respondents

Court File No. 211/19

ONTARIO
SUPERIOR COURT OF JUSTICE
(DIVISIONAL COURT)

PROCEEDING COMMENCED AT
TORONTO

AFFIDAVIT OF BEN GREEN
SWORN MAY , 2019

FOGLER, RUBINOFF LLP
Lawyers
77 King Street West
Suite 3000, P.O. Box 95
TD Centre North Tower
Toronto, ON M5K 1G8

Young Park (LSO# 43550E)
ypark@foglers.com

Tel: 416.365.3727
Fax: 416.941.8852

Samantha M. Green (LSO# 63680N)
sgreen@foglers.com

Tel: 416.860.6904
Fax: 416.941.8852

Robert B. Macdonald (LSO# 60512B)
rmacdonald@foglers.com

Tel: 647.729.0754
Fax: 416.941.8852

Lawyers for the Applicants