

## EU CSS Symposium – Online vs offline inequality: Examining disableist infrastructure via open data.

Vanessa Thomas, 16 Nov 2017

In the spring of 2017, I was challenged to create a cool “open data science” project. I had grown pretty bored of seeing Twitter and Wikipedia projects—no offense to anyone here working on twitter or Wikipedia projects—so I set out to find what I considered to be an interesting open data set released by **some** city government, **somewhere**.

I’m Canadian, so the first place I decided to check was Canadian cities, and I started with the City of Vancouver’s open data repository, which has a really positive reputation within Canada.

As I was scrolling through their list of datasets, one immediately caught my attention:

A public washrooms data set!!

I opened it up, saw that that it included data points for 105 public washrooms around the city, AND that it included information about the washrooms’ wheelchair accessibility status. And I immediately decided that I could do something interesting and maybe even useful with this information.

The project that unfolded—and that I continue to work on—from that data set is what I’m here to talk about today. I’ll be talking about toilets. I’ll be talking about open data.

And I’ll be talking about how digital technologies—and computational social science—can expose, reproduce, and transpose inequality, especially in this case infrastructural inequality related to disability. But that all needs a bit of unpacking. And it’d probably help for me to briefly introduce myself before I do that, so that you can get a sense for where I’m coming from with this project.

I’m Vanessa Thomas, a queer Canadian postdoctoral research fellow with the Department of Computer Science at Aarhus University in Denmark. I have an intensely interdisciplinary background that includes everything from advanced calculus, software development, and systems architecture design to public policy, science communication, human rights, and peacebuilding studies.

I'm going to draw on my multidisciplinary background to talk with you today about "online vs offline inequality"—one of the themes proposed in this symposium's original call for papers.

To get to my main points about online vs offline inequality, I need to tell you a bit more about the data set I chose to work with and what I've done with it.

So. Vancouver's public washroom data set is quite unique; very few major cities publish open data related to their public washroom infrastructure, and even fewer include wheelchair accessibility information within those datasets. In fact, no other Canadian city provides information about wheelchair accessibility in their "public washrooms" data sets.

In the City of Vancouver's dataset, they provide information about the name, location, hours, and wheelchair accessibility of 105 public washrooms in XLS, CSV, and KML file formats. And according to the version that I encountered in early 2017, merely fifteen of Vancouver's 105 public washrooms are wheelchair accessible, just over fourteen percent. For those of you who don't know, Vancouver is renowned for its wheelchair accessibility, so its low percentage of accessible toilets struck me as incredibly surprising.

In fact, I found it so surprising that I decided to visit all 105 washrooms during a trip to Vancouver in July 2017. I spent five days cycling around Vancouver, visiting and verifying the accessibility of every single toilet in the dataset. I also spoke with two occupational therapists about the infrastructural requirements for wheelchair accessible toilets—things like door width, arm supports, accessible sinks, garbage bin placement—just in case I wanted to take any measurements or document specific details about the facilities I encountered. And then I cycled. And I cycled. And I cycled. To all 105 toilets.

By the time I completed my verification process, I had identified over two dozen issues with the dataset. I found that at least fifteen wheelchair accessible toilets that had been mislabelled as inaccessible, two inaccessible toilets were labelled as being accessible, eleven appeared as if they might have met the criteria for accessible toilets, and numerous community centres (with public washrooms) were inexplicably excluded from the dataset. I've been in touch with the city of Vancouver about this, they had some interesting reasons for the inaccuracies, and they're working to fix them.

Beyond these concrete, long-term data validity issues that can easily be addressed in the data set, I also encountered numerous on-the-ground ephemeral issues. For example, in these three images: on the left, this washroom was inexplicably closed,

with no reason posted as to why or for how long. In the top right, the placement of the garbage bin likely prevented easy access to the toilet if you were using a wheelchair. In the bottom, the actual toilets were closed, and a bunch of porta-potties were placed haphazardly outside. In other cases, one accessible toilet was temporarily closed for cleaning, another was closed due to movie filming and another was temporarily inaccessible due to a water main break.

So. As you can see, there are considerable issues with the dataset. This mattered to me because I had originally planned to use the data set to directly and critically examine how much more difficult it was to travel to wheelchair accessible vs inaccessible infrastructure in Vancouver. If I hadn't thought to verify the dataset in-person, I would be conducting analysis using already problematic data.

Which is why the ephemeral and enduring data validity issues I've mentioned underscore how (online) open data—and any calculations or projects based on them—can expose, reproduce, and transpose (offline) inequalities, especially in this infrastructural, mobility-related disability case. So what do I mean by “expose, reproduce, and transpose” inequalities?

In this specific case, I mean that I can use this open data to expose, through visualisation, the gaps in Vancouver's municipal infrastructure for wheelchair users.

If we look at the map again, we can see spots where there are no accessible toilets and we can expose the lack of infrastructural support for an already marginalised community.

Of course, I can also use the original, unverified data set to reproduce inequality, and in this case disableist inequality, by providing misleading or inaccurate data that suggests wheelchair users are even more underserved than they actually are in Vancouver.

I'll pause here quickly because disabelism might not be a term that everyone is familiar with, so to clarify: I'm using the terms disableism and disableist in alignment with Gregor Wolbring's definitions, referring to anything that discriminates against the 'less able'.

Lastly, I can also use the original, unverified open data set to transpose disableist inequality from being an (offline) infrastructural deficit into an (online) informational deficits, which can then also influence how, where, and when people engage (offline) with Vancouver's urban infrastructure. And it can influence how we, as researchers, are perceived for conducting research using inaccurate data.

So, for example, in this case, if the dataset—or academic analysis on that dataset—says a wheelchair accessible toilet exists, but someone shows up and it doesn't or it's inexplicably and temporarily closed, they are less likely to trust the other information in the dataset, regardless of its accuracy.

Now we get to the so what part of the presentation. Why does this matter to computational social science? Well, I believe that my project thus far—and the fact that it happened to be me (a visiting, white, female, middle-class, able-bodied, computing researcher who knew enough about Vancouver's accessibility reputation to be surprised by its dataset, and who also had enough spare time to cycle to all of the city's public washrooms) who identified issues with this dataset—raises fundamental questions for computational social science research built on open data.

For example, what role(s) do computational social science projects (unknowingly?) play in perpetuating (offline) inequalities when those projects rely on (online) open data? How can we navigate and negotiate organisational issues, like a lack of funding or personnel, that influence the validity of data? How can we use our positions of privilege as computational social science researchers to challenge inequalities that are perpetuated by data? What would emancipatory computational social science *look* like? Is such a thing even possible?

I don't have answers to these questions. I'm hoping some of you might have answers or ideas. My reviewers mentioned crowdsourcing as a potential solution to some of the issues I raised, but crowdsourcing brings its own challenges and won't address the data maintenance that currently needs to take place in the City of Vancouver's administrative offices. I could tell you plenty of other stories, but they could keep me talking for hours.

But anyway, since I don't have any answers to my questions, I'll leave you with what I do have, which is a few more toilet photos.

Thank you for your time, thank you for listening. I'll turn it back over to you for questions!