

Voici quelques exemples de données que nous ignorons :

- Le nombre de personnes vivant dans des sous-locations illégales à New York ;
- La traçabilité des armes vendues aux États-Unis et les informations sur leurs propriétaires ;
- Le nombre de gens expulsés des États-Unis selon l'État où ils vivaient ;
- Le nombre de Rohingyas en Birmanie.

J'emploie le terme « Missing Datasets », pour désigner ces informations manquantes, angles morts d'un monde qui semble aujourd'hui criblé de données. Autant de fantômes qui font penser à la liste scotchée sur un coin de mon bureau. Ces données représentent tout autant la réalité de notre monde, summum d'une quantification qui a été mise de côté. Nous ne mesurons que les choses qui nous intéressent suffisamment. Les données manquantes ont aussi de la valeur, mais elles ne peuvent pas être mesurées.

Mon recueil des données manquantes existe sous des formes bien plus pérennes qu'un bout de papier. Il prend notamment la forme d'une œuvre d'art intitulée *The Library of Missing Datasets*. Au premier abord, cette « bibliothèque des données ignorées » ressemble à un banal bloc de tiroirs, mais les étiquettes des classeurs qu'il renferme identifient ces absences. Les classeurs sont évidemment vides, puisque les données sont inexistantes.

Je me suis faite la documentaliste de cette bibliothèque qui ne cesse de s'agrandir. À travers ces absences, j'ai découvert que l'exclusion suit des modèles précis, qu'il existe des structures qui décident de ce qui doit ou ne doit pas être inventorié. J'ai pris note des caractéristiques qui font que certains endroits sont immunisés contre cette quantification croissante du monde. Plus d'une fois, je me suis retrouvée à aider des gens pour recueillir des données auparavant manquantes, ou bien à expliquer pourquoi tout ne pouvait pas — ou ne devait pas — être mesuré.

Plus la liste s'allonge, plus je suis frappée par les questions hautement symboliques que soulèvent ces données fantômes. Elles ne risquent pas de disparaître : tant que nous catégorisons les choses et que nous cataloguons le monde en conséquence, il y aura toujours des données manquant à l'appel. Il y aura toujours des informations qui échapperont aux tableaux de chiffres, des choses qu'on ne peut pas ou qu'on ne doit pas prendre en compte. L'appréhension du monde par la discrimination implique une certaine simplicité, et les données manquantes, en vertu de leur existence et de leur absence, remettent en cause cette simplification.

La complexité et le désordre de ces données sont passionnants, car elles trahissent un certain type de pouvoir. Une absence toujours remarquée laisse entrevoir le spectre d'un autre monde, où les priorités seraient différentes. Aucune donnée n'existe sur les violences policières faites aux Amérindiens, mais que se passerait-il si c'était le cas ?

Ces données manquantes n'apportent aucune réponse, mais elles font office de rappel cinglant : nous sommes responsables de la manière dont nous cataloguons le monde. En choisissant les données à prendre en compte et en leur allouant une crédibilité, nous déterminons les limites de notre monde. Si tel est le cas, alors nous sommes aussi capables de changer cet état de fait, et à chaque instant, de changer notre monde.

« Plus la liste s'allonge, plus je suis frappée par les questions hautement symboliques que soulèvent ces données fantômes. »

What Is Missing Is Still There



My interest in missing things began with what I could see. For a long time, I have kept a small piece of paper taped to the bottom right corner of my desk. This paper comes and goes, at times becoming wrinkled, discolored by tea stains, or hidden under a stack of books. But it always serves the same purpose: listing the most eccentric datasets that I can find online.

Before the score and lyrics for the hit American musical *Hamilton* had been released, a group of obsessed fans created a shared document of every word in the show. This dataset made my list. In 2016, a Reddit user published a post with a link to where he had downloaded the metadata of every story ever published on fanfiction.net, a popular site for stories about fandoms. This, too made the list.

Other things that have graced the list: the daily count of footballs produced by the Wilson Sporting Goods football factory in Ada, Iowa (4,000, as of 2008); an estimation of the number of hot dogs eaten by Americans on the 4th of July every year—most recently 150 million; the locations of every public toilet in Australia (of which there are more than 17,000).

Australian academic Mitchell Whitelaw defines data as measurements extracted from the flux of the real. When we typically think of collecting data, we think of big, important things: census information, UN data about health and diseases, data mined by large companies like Google, Amazon, or Facebook.

From this perspective, Whitelaw's definition of data is admirably concise and effective. With its clever use of the word "extraction", it hints at the resource-driven nature of data collection. Like Shoshana Zuboff's concept of surveillance capitalism, which describes our modern ascendance into a form of capitalism that monetizes data gathered through perfunctory surveillance, Whitelaw's definition calls to mind corporate imaginings of data as a resource. In a capitalist society, it is always a smart business decision to collect data. A world collected is a world classified is a world rendered legible is a world made profitable.

But when I glance at the list on my desk, it is not always easy to spot the direct line that connects them to these concepts of resource-extraction and omnipresent surveillance. While less conventional, the data sets on that list also are vertices of quantification, facts extracted from surprising corners of reality. And so a simpler definition comes to mind.

Data: the things that we measure and care about.

This is the beauty I find in the list of odd data on my desk. If Whitelaw's definition suggests a world that is pure source, a heap of raw material waiting to be cut up and structured into neat cells and excel spreadsheets, then mine highlights the opposite: the fact that all datasets are created by people who have a stake in their creation.

The corollary is also true. If we wish to know more about what our societies, corporations, and communities value, we should simply look to what data is collected. The things we measure are the things we care about.

When I first began creating my list of weird data, I wasn't sure why I was doing it. Idle curiosity seemed the most obvious reason, and fascination with novel forms of procrastination another.

But at some point, the answer became clear to me. When it did, I added an additional item to the piece of paper. This item was a quote, taken from an old conversation that I had had with a former colleague.

"And as the list grows, I have increasingly been struck by the symbolic questions these shadow datasets raise."

"Humans make sense of the world through exclusion."

The quote came from John Fass, a fellow researcher from the Royal College of Art who had focused on design and interfaces. John and I had been talking in the empty canteen one day when he offhandedly mentioned that he considered exclusion to be a crucial aspect of design. The only way that humans were able to make sense of the world, he insisted, was by sifting through information and making decisions about what needed to be excluded at any given time. Narratives only work because of the many mundane details that are removed in the course of their telling. In a sense, all stories we tell ourselves are exercises in leaving things out.

It was not the first time I had heard this concept, but on that day it resonated with me. In their seminal (and very dry) academic text, *Sorting Things Out*, Geoffrey Bowker and Susan Leigh Star title the book's introduction with the phrase, "To classify is human". They argue that our understanding of the world depends on the use and creation of implicit categories that serve to order the world. The difference between outdoors and indoors, for instance, dictates different ways of dress, activities, and so on.

But later on, Bowker and Star push a more incisive point about classification. "No one classification system organizes reality for everyone", they warn. "For example, the red light, yellow light, green light traffic light distinctions do not work for blind people (who need sound coding). In looking to classification schemes as ways of ordering the past, it is easy to forget those who have been overlooked in this way."

Datasets are the end products of classification systems, the clean outputs of intentional orderings. My list of odd datasets was just the tiniest gesture at the many ways in which we have thought to classify our world.

But the same way that a traffic light shows what we prioritize (vision) and cannot work for everyone (the blind), datasets point to their own contrasts—specifically the things that we haven't collected. And if it is true that we make sense of the world through exclusion, then perhaps there is a special type of meaning to be found in the things that we leave out.

Here are examples of some of things we do not know:

- The number of people living off-lease, in illegal housing situations in New York City;
- Gun trace data for people in the US who have bought guns;
- Which state people deported from the US were living in at the time of their removal;
- The number of Rohingya people in Myanmar.

Missing Datasets is the term I have for these blank spots in a world that nowadays seems soaked in data. They form a ghostly parallel to the sheet of paper that occasionally adorns my desk. They, too, are the facts of our world, the vertices of measurements. But they are the ones that we know little about. Data are what people care about enough to measure. Missing datasets are the things that people care about, but cannot measure.

My repository of Missing Datasets lives in forms far more permanent than a sheet of paper. One of these forms is an art piece called *The Library of Missing Datasets*. On first glance, it appears as simply a painted filing cabinet. But it holds within its drawers physical folders upon whose tabs the title of a missing dataset has been inscribed. The folders are empty. The content, like the data, is missing.

I've made myself a shepherd over this ever-growing library of missing datasets. Through them, I've learned that there are patterns to exclusion, structures that govern what is and isn't able to be collected. I've taken note of the characteristics that make places immune to the growing datafication of the world. More than once I have found myself helping a group to collect some data that once was missing, or justifying to another why not everything can or should be collected.

And as the list grows, I have increasingly been struck by the symbolic questions these shadow datasets raise. Their existence is assured: as long as we classify things and sort the world according to these classifications, there will always be missing datasets. There will always be bits that ooze out beneath spreadsheet cells, things that cannot be contained, or that should not. Making sense of the world through exclusion implies a certain simplicity, and missing datasets, by virtue of their existence and nonexistence, challenge that simplicity.

I find this difficulty and its messiness thrilling, for it betrays a type of power. If something always is missing, it means that we have always the spectre of a different kind of world, with different kinds of priorities. We do not collect data on police violence against Native Americans—but what kind of world would it be if we did?

These missing datasets do not provide answers, but the reminders they carry are poignant. We are the ones who render this world collectible. Each time we choose what data to collect and imbue that data with validity, we define the terms of the world. But if so, then we are also the ones capable of changing it, and making it different, each and every time.

"I find this difficulty and its messiness thrilling, for it betrays a type of power."