Good Technology is Slow (to Scale)

Ranjit Singh
Researcher at the AI on the Ground Initiative of Data & Society Research Institute

---

We met Ranjit through one of Eleanor's friends, Michelle Spektor, an expert in the history of biometrics. Ranjit explores concepts, keywords, and everyday stories about living with data and AI in and from the Majority World (an alternative to the terms 'developing world' or 'Third World' that describes the countries where the majority of the world's population resides). This covers everything from experiences of biometric surveillance systems like the Aadhar card system in India to creating resources and primers about AI in and from the Majority World. Ranjit was a joyful presence on the podcast and also brings his trademark creativity, warmth and enthusiasm to his academic work. He introduced to us the idea of biometric systems 'seeing' in 'low and high resolution', foregrounding some people and aspects of the human experience and obfuscating others. These are not 'just' metaphors; we rely on these kinds of ideas to help us understand the true impact of identification systems in shaping our lives and the allocation of state resources.

Slow is a culture. It is a movement. It is a way of life. Slow is a reaction to the increasing pace of life in the fast lane, from fast food to fast scholarship. As Carl Honore[1], one of the prime proponents of the slow movement, puts it: Slow is about taking the time to do something as well as possible, not as fast as possible. Slow is a response to efficiency.

This piece is not in praise of slowness; it is an invitation to it. It connects slowness with scale, particularly in response to quickness of data systems that have come to shape our everyday lives. It explores scale in relationship to two prominent features of big data systems: volume and velocity. Volume refers to quantity of accumulated data; big volume implies large quantities of data that cannot simply be read by humans. We need machines, specifically computers, to understand it. Velocity marks the speed at which data is accumulated. The greater the velocity, the faster the accumulation, the more difficult it becomes to process data and make sense of it.

In exploring a dozen ways to get lost in large scale data sets, Lawrence Busch[2] begins with a simple distinction of scales in managing volume of data. Scale in big data operates at two levels: aggregation and individuation. Aggregation involves combining individual data (records or even, datasets) to create large volumes of data. These large volumes of data are processed to find correlations that connect groups with certain characteristic patterns of behavior. Aggregation produces patterns in similarities; scale becomes a matter of managing the largeness of volumes of data. Individuation involves establishing characteristics of an individual user in relation to patterns of aggregated group behavior. The individual only exists as unique in relationship to these aggregates. Individuation produces patterns through differences; scale becomes a matter of managing the smallness of individual datasets.

Scale in managing volume is a matter of establishing relationships between data records. What does the data represent collectively is a question of aggregation. How does the data represent an individual's place and position in a dataset is a question of individuation.

Scale not only represents but also creates these relationships. In representing aggregate behavior of people, it creates people who fall into categories of "desirable" behavior. 'Trustworthy' people, 'Punctual' people, 'Unique' people, this list can go on. These categories of 'desirable' behavior inevitably come with their inverse.[3] 'Trustworthy' people can only exist in relation to 'untrustworthy' people. Otherwise, everyone is trustworthy, and no one is trustworthy. People not only have to fit into these new categories of 'desirable' behavior, but they also must know when they fall into its inverse and work their way out of it.

Now imagine if this happens quickly. One day you're 'trustworthy', the next day you're not. With increasing use of data systems in all aspects of everyday life, these decisions based on sociohistorical datasets and made by machines have had real consequences at a much faster pace. People are left to face these uneven consequences, often without an explanation about how such decisions are made.

Examples of such consequences include the lawsuit in Kenya around double registration of vulnerable citizens[4], who have struggled to obtain citizenship documents because their biometrics were recorded in refugee databases. One day you're a citizen, the next day you're a refugee. And the class action lawsuit in Australia against RoboDebt[5], an automated debt recovery scheme, which wrongly accused about 400,000 welfare recipients of misreporting their income and issued reams of debt notices. One day you're a welfare beneficiary, the next day you're asked to return your welfare payments.

This brings me to the first point that I wish to make with this piece: The pace of such decisions must become slower, especially when there are life chances at stake. This slowness requires erring on the side of generosity and inclusion, rather than efficiency and exclusion.

During my fieldwork on the challenges of appropriating India's biometrics-based identity number in welfare services[6], a respondent narrated a parable on unanticipated consequences of interventions. It provides a window into this call for erring on the side of generosity.

> There once was a man who pledged that he would feed the pigeons that used to gather on his porch every day. Over time, he realized that there were not only pigeons eating the feed, but also some crows. The moral of this story lies in one simple question: Should this man stop feeding the pigeons because of the crows? And, I believe, the answer should be, "No" (Fieldnotes, 2 July 2015).

Given the inevitability of failures associated with any data system, the work of maintaining "good" data systems must err on the side of generosity.

Erring on the side of generosity involves four crucial steps:

- First is to start small and with multiple pilot projects across places and contexts. These projects must be complemented with existing processes that provide alternative ways to achieve the same goals when data systems fail. These pilot projects offer an opportunity to iterate on experimenting with how well a data system works and its consequences across scales of place and time. These systems can be connected with each other over time to produce data infrastructures.
- Second, ensuring that data systems are incorporated in everyday life linearly (or at a steady rate), rather than exponentially (or at a compound rate). Societal change is slow and tends to unfold linearly, while data systems are fast and grow exponentially. The disconnect between the two often produces breakdowns in how data systems are incorporated in any social process.
- Third, moving at the speed of trust between system operators and people who will eventually become subject to the system for accessing social services. People need time to develop their own understanding and literacy in interacting with these systems. They need opportunities to speak back and seek due process when they face harm.
- Fourth, accepting inefficiencies and errors as a part of the operating cost of the system, rather than passing on this cost on to the people. This happens most often in determining eligibility of a person in receiving a service, especially in the

> context of targeted welfare services such as RoboDebt.[7] The parable was also in response to concerns around determining eligibility. An automated decision of ineligibility or debt notice (in RoboDebt's case) must not automatically exclude an existing beneficiary from a welfare service without due process.

All these steps take time. They require that we slowdown in building and living with data systems.

This brings me to the second point I wish to make with this piece: slowing down has deep implications for managing velocity of data. This velocity spans from the pace of data collection, processing, and analysis to the quickness of data-driven decision-making. Velocity is often treated as synonymous with efficiency in the discourse of big data.

Its value has often been illustrated with a story of selling Halloween cookies at Walmart. This story, as documented by Bernard Marr on real-time insights from Walmart's data cloud in 2017[8], goes as follows:

> Sales data showed that a particular kind of novelty cookie launched to celebrate Halloween was very popular in most Walmart stores, except two where they were not selling at all. This difference was quickly investigated. The investigation revealed that a simple stocking oversight had resulted in the cookies not being placed on the shelves at all in the two stores. The cookies were immediately put on shelves preventing further loss of sales.

The story represents insights emerging from both aggregation and individuation. While the Halloween cookies were very successful as an *aggregate* pattern in sales data, there were two stores that *individually* were not doing as well as expected in terms of the aggregated pattern. This story is about not just the speed of data analysis and interpretation, but also the speed of making real-time decisions based on this data. Had the stocking oversight been noticed after Halloween, the cookies may not have sold at all.

Certain decisions are time sensitive. They range from simple logistical issues at supermarkets to humanitarian crisis responses. Responding quickly is essential, even if it is with incomplete data. Doing something at pace is better than waiting in hopes of doing it better. However, not every data-driven decision is time-sensitive. Particularly, when it comes to determining life chances, decisions to exclude must not be taken without due process.

Making data-driven judgements on citizenship status in Kenya and welfare eligibility in Australia are examples of such crucial decisions that need time. When decisions are made quickly, they become prone to bias, which often stacks the odds against the systemically marginalized. Slowing down would inevitably involve understanding the unevenness of bias and finding ways to tackle it.

Scale as a matter of managing velocity represents scales of time. This is obvious in every call for due process that demands that time taken to make a decision must correspond with the magnitude of consequences and harms of incorrect decisions for people who must live with that decision. If the potential of harm is greater, the time taken for due process before deciding must be correspondingly longer.

These calls, however, tend to take the infrastructure required for data systems to produce automated decisions as a given. When we open the process of building this infrastructure to scrutiny, we will find that all infrastructural change is slow, whether it is building a new subway line in a city or creating the conditions for the use of data systems in delivering any service. Automated decision making based on data has tremendous infrastructural momentum. This momentum is a result of large-scale investments of resources made into building data infrastructures and producing volumes of data to inform decision making. These investments have now become justifications for appropriating data systems in existing practices of providing services. The impetus on velocity comes from these justifications and resulting expectations of quick returns from ongoing investments.

However, existing practices do not simply roll over and change when faced with such infrastructural momentum. They often exhibit infrastructural inertia. Infrastructural inertia does not imply stasis; rather it is the work required to change existing practices and develop new competencies in transitioning to new processes of data-driven decision making. The slowness in infrastructuring data-driven decision making emerges in the mutual shaping of this momentum of data systems and inertia of existing practices.

It is during this slow process of mutual shaping that meaningful interventions in erring on the side of generosity can be made. Rather than reacting to the harms of data-driven decision-making, it is time that we proactively account for its consequences and prepare for them. Whether this accounting happens through algorithmic impact assessments or through algorithmic audits, its effect is to slow the momentum of data-driven decision making so that its consequences can be assessed/audited before its deployment in the real world.

Let us strengthen the inertia of transitioning to data-driven decision making. This is not resistance for its own sake; it is the only way of ensuring that the dignity of data subjects is not the cost of efficiency. Being slow does not mean rejecting change. It means embracing change thoughtfully. A technology that enables us to do this…that scales slowly and thoughtfully is 'good.'

---

[1] Carl Honore, *In Praise of Slowness: Challenging the Cult of Speed*, Annotated edition (New York, NY: HarperOne, 2005).

[2] Lawrence Busch, "Big Data, Big Questions | A Dozen Ways to Get Lost in Translation: Inherent Challenges in Large Scale Data Sets," *International Journal of Communication* 8, no. 1727–1744 (June 2014), http://ijoc.org/index.php/ijoc/article/view/2160/1160.

[3] Ranjit Singh, "'The Living Dead': Orphaning in Aadhaar-Enabled Distribution of Welfare Pensions in Rajasthan," *Public* 30, no. 60 (2020): 92–104, https://doi.org/10.1386/public_00008_7.

[4] Haki na Sheria Initiative, "Biometric Purgatory: How the Double Registration of Vulnerable Kenyan Citizens in the UNHCR Database Left Them at Risk of Statelessness," Citizenship Rights in Africa Initiative, November 17, 2021, https://citizenshiprightsafrica.org/biometric-purgatory-how-the-double-registration-of-vulnerable-kenyan-citizens-in-the-unhcr-database-left-them-at-risk-of-statelessness/.

[5] Terry Carney, "Robo-Debt Illegality: The Seven Veils of Failed Guarantees of the Rule of Law?," *Alternative Law Journal* 44, no. 1 (March 1, 2019): 4–10, https://doi.org/10.1177/1037969X18815913.

[6] Ranjit Singh and Steven Jackson, "Seeing Like an Infrastructure: Low-Resolution Citizens and the Aadhaar Identification Project," *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW2 (October 18, 2021): 315:1-315:26, https://doi.org/10.1145/3476056.

[7] Lyndal N. Sleep, "Trauma and Automated Welfare Compliance in Australia," Points | D&S Research Blog, December 14, 2022, https://points.datasociety.net/trauma-and-automated-welfare-compliance-in-australia-ba661a60b50f.

[8] Bernard Marr, "Really Big Data At Walmart: Real-Time Insights From Their 40+ Petabyte Data Cloud," Forbes, accessed November 22, 2022, https://www.forbes.com/sites/bernardmarr/2017/01/23/really-big-data-at-walmart-real-time-insights-from-their-40-petabyte-data-cloud/.