**Building the Framework of the Politics of Numbers**

**My big “aha” moment from D’Ignazio and Klein 2020 that motivates this work:**

Expose the *privilege hazard*: “[to examine power in and around data]...often means asking uncomfortable questions: who is doing the work of data science (and who is not)? Whose goals are prioritized in data science (and whose are not)? And who benefits from data science (and who is either overlooked or actively harmed)?”

**Part I: The Promise**

**Q1: Why do we want big quantified/statistical data in development work? What is the promise of development data?**

* “Make the situation of the poor visible to policy makers” (Serajuddin et al 2015, 2)
* “What gets counted, counts” - when statistics make something visible, it can inform policy making and resources allocation. “Being represented means being made visible” (D’Ignazio and Klein 2020, 110). “Quantification is representation” (39) / Generalizable
* Can help diagnosis - more refined analysis illuminates more targeted areas
* Data needed to keep track of progress towards goals - a form of accountability
* Quantified data may be perceived as more objective and persuasive; more robust/less biases (less anecdotal)
* Quantitative data *appears* objective/less biased, can be more powerful/persuasive to \*certain\* audiences, reveal patterns.
* Cover your Ass (CYA) strategies? Gives legitimacy to policymakers’ decisions. Stats can help motivate and bolster decisions. Justification.
* Stats can often help people better understand problems.
* Also needed for mobilization of resources and for program design ex: there is not research/info on how to properly take trans folks experiences into account post disaster, so we still have people unable to access aid because their drivers license/ID might not match their gender expression
* Data as means of learning and revealing:
  + When data gets re-examined through different means of introducing different strata or groups, disaggregation, classification, etc… can reveals inequalities and patterns of bias or discrimination that we were not previously aware of.
  + Can reveal discrepancies, refine research, virtuous cycle of big data - reflection - big data…
  + COVID infection and mortality rates: what happens when we really disaggregate and map the data?
  + Ex: what happens when we add new gender categories or options in data collection and reporting?
  + Quantified data can reveal system level oppression, especially around issues that are not generally talked about
  + D’Ignazio and Klein (2020): women don’t talk about their experience with maternal healthcare, may think it “just happened to them” or its an anomaly until we start to see patterns revealed through systematic data collection. (other examples related to maternal health: miscarriages, difficulties breastfeeding).
* Data as tool of accountability
  + Data does not just inform policy, it informs public opinion and can mobilize pressure/action
  + Data needed to keep track of progress towards goals - a form of accountability

**Q2: What benefits does a data revolution provide us in development work? Where has progress in data collection and dissemination revealed new dimensions of poverty and development that have improved (in your opinion) the way we think about and pursue our work? *(provide here specific examples of promises presented by the data revolution*)**

Disaggregated data (both intentionality - we know know to collect data in this form

Granular data can be a tool of accountability - e.g. measuring retirement of coal, natural gas reliance / consumption data

More incorporation of mixed methods to improve causal analysis

Improvements in frequency of data collection (e.g. poverty data, HHS, DHS) - more systematic data collection

More data available means more people can get involved in analysis, audits, idea generation, testing of theories,

E.g. cash transfer studies, helping us better bridge gap between different programs and delve more into informal economy

E.g. FEWS (allows for timely warning around natural disaster situations - getting information out quickly). Broader revolution in terms of technology - e.g. cell phones

Big data can look at secondary effects (by pairing with other big data) - “potential for mash-ups”

Causal inference!

We can often see what data is NOT being collected (e.f. Femicide maps in Mexico)

Find information that is otherwise hard to track. E.g. climate migration data - are we seeing IDPs or refugees? What is the geography of forced migration?

**Q3: What are the arguments in favor of establishing major international development goals, such as the SDGs, with quantified targets and indicators?**

**Q4: What are the perils of the data for development revolution and the focus on quantifiable, statistical data?**

***General Arguments:***

1. What doesn’t get counted, doesn’t count . Are we conuting what we want to count, or what is possible/easy to count?
   1. E.g. Maternal mortality of black women
2. Measurement often captures outcomes, but not causes
3. The Rise of the Counterbureaucracy and the ‘Obsessive Measurement Disorder” (Natios)

*Counter-bureaucracy* - “a set of US agencies charged with command and control of the federal bureaucracy through a set of budgeting, oversight, accountability, and measurement systems that have grown over several decades to a massive degree, with extraordinary layer upon layer of procedural and compliance requirements.” (Natsios 2010, 3)

* 1. “Ignores the central principle of development theory - that those development programs that are most precisely and easily measured are the least transformational, and those programs that are most transformational are the least measurable.” (p.3)
  2. Ie, we have moved into system were we pursue what is easily counted, rather than what really counts, and within timelines that are infeasible for showing actual outcomes or impact of development work. (p7). Has created dangerous risk aversion
  3. “...measurability should not be confused with development significance” (p.9)
  4. We should not confuse compliance or accountability with development effectiveness” (Natsios 2010, 37)

1. Obsessive Measurement Disorder: “an intellectual dysfunction rooted in the notion that counting everything in government programs (or private industry and increasingly some foundations) will produce better policy choices and improved management.” (p.3)
   1. This drains time and human resource away from “real’ development work, both for donor staff and recipient country partners.
   2. E.g. Bright magazine article - waiting for data to show up when you already know the problem exists

**Part II: The Perils**

**Q5: What do we understand to be the general risks of creating rules and cultures around data-driven development?**

1. ***Data deprivations (Serajuddin et al 2015)***
   1. coverage,time gaps (frequency), sampling errors, measurement errors
2. ***Empirical challenge of field data collection (underlying cause of data deprivations)***
   1. Frequency, coverage, sampling
      1. Difficulties in conducting HHSs - a “bottleneck for poverty data” (p.3)
      2. Difficulties in reconciling survey vs administrative data
      3. Non-comparability over time as debates over concepts and constructs lead to changes in what is measured and how
      4. Costs of critical HHS (including LSMS, DHS, Labor Force Surveys, Multiple Indicator Cluster Survey or MICS)
      5. E.g. maternal mortality data (and other “rare event” data)
      6. Food insecurity (esp, in conflict zones)
      7. Migrant/ nomadic/ unregistered populations
3. ***Classifications***
   1. Overly broad classifications can create false options/dichotomies
   2. What are the perils of too narrowly defining groups? (the further you break down strata, the more likely they will be so small in the final outcomes wrt data points that they disappear in the analysis).
   3. Census data: Congressional Black Caucus and civil rights group opposition to creating more / multiracial categories / giving choice to click multiple racial categories) could reduce their representative voice (D’Ignazio and Klein 2020, 104).
4. ***Complexity of Constructing Composite Indicators (construct and content validity)***
5. ***Statistical Capacity Challenges***
   1. NSO capacity: Jerven 2013
   2. National funding vs donor-drive agendas (see also Sandefur and Glassman 2015, 4)
   3. Civil Society limited capacity to help collect/use data
   4. Digital divides: big AI technology is benefiting some, but leaving others in the dust

***6. Open Data*** - risk of garbage in, garbage out (Sandefur and Glassman 2015, 24); open data without critical data skills or false assumption that “if you build it they will come”? Naive optimism about capacity and willingness to use open data for good.

***7. AI technology - getting the software right (and the dangers of privilege hazards, D’Ignazio and Klein 2020)***

* 1. Profiling technology, when based on skewed biometrics (e.g. white man for car crash tests)
  2. Name databases (e.g. TSA potential terrorist list)
  3. Technology that relies on human decisions (TSA - gender assumptions/x-ray scans)

**Q: What are some of measurement or empirical challenges we face in collecting and using high quality, quantifiable data?**

**Q: What are some of the political challenges we face with respect to data collection, reporting, and use?**

**Part III: How Do We Maximize the Promise and Minimize the Perils of a Data for Development Revolution?**