



Dear Chair Koo, Acting Public Advocate Johnson, and members of the committee and commission:

Our world has been transformed over the past twenty years. We're living in a new Industrial Revolution. Data are automating the way in which we do everything – live, work, shop and play. The private sector uses data to figure out what consumers want now, and uses data to deliver to them – with almost frightening efficiency.

Except government agencies that produce data for decision-making. Those agencies are often doing their job in the data dark ages – producing information manually, and using dated tools and methods to do so. But governments can't simply turn to the private sector to do their job for them. Government data are used to allocate millions of taxpayer dollars, affect millions of citizens, and are intended to serve the public, not generate profit. Governments need to be trustworthy, generate information that are measured well and consistently over time and need to ensure that confidential information is protected.

The bad news is that the way governments produce data won't change by itself. In the private sector, market forces create the impetus for change, because organizations that don't adapt get driven out of business. There's no similar force driving government change. Over the past 30 years, I've worked with people at all levels of government – federal, state, county and city – both within the US and in the rest of the world. I've developed tremendous respect and admiration for the highly skilled and dedicated workforce that brings us the information that drives our economy. They know what needs to be done to make change happen. The City Council's focus on data and transparency offers a tremendous opportunity to effect change if well designed and well implemented.

My name is Julia Lane, and I am a Professor at New York University's Robert F. Wagner Graduate School of Public Service. I have worked with government data for my entire career. I cofounded the US Census Bureau's Longitudinal Employer Household Dynamics Program, based on state level unemployment insurance wage record data, which is the first, and still the only US statistical program established by a researcher. I founded the first remote access secure data enclave for government data at the National Opinion Research Data Center at the University of Chicago. I also established the NYU Coleridge Initiative, which created the Administrative Data Research Facility as well as the Applied Data Analytics training programs. I am an elected fellow of the American Association for the Advancement of Science, the American Statistical Association, and the International Statistical Institute. I have received the Roger Herriot Award, the Julius Shiskin Award and the Warren E. Miller award for my services to statistics and social science.

An effort to focus on data and transparency could effect change if it ensures that the right technology is put in place to protect the confidentiality of data, that agencies are engaged in the effort to change the way in which they do business, and that their staff are empowered and trained to make use of data with the appropriate methods and tools. I expand on these points in my remarks below.



**First**, the technology exists to enable agencies to share data securely and responsibly. Investments have been made at the Federal level to support the decision making of the Commission on Evidence-Based Policymaking. Ten of the Commission’s recommendations have been signed into law: H.R. 4174, the “Foundations for Evidence-Based Policymaking Act of 2018,” which improves evidence-based policy through strengthening Federal agency evaluation capacity; furthering interagency data sharing and open data efforts; and improving access to data for statistical purposes while protecting confidential information. Much has been learned from important national investments in the area of Secure and Trustworthy Cyberspace as well as the National Privacy Research Strategy. A great deal has been learned about risk management, ensuring that information flows and use are consistent with privacy rules, and the human aspects of building secure environment. Useful strategies that exist to improve computer and cybersecurity and there has been an important debate about how best to securely release aggregate information, particularly whether new cryptographic techniques, such as differential privacy, can be broadly applied.

The NYU Coleridge Initiative’s Administrative Data Research Facility (ADRF), which was established less than three years ago to inform the Commission’s decision making, already provides a secure environment within which many agencies from jurisdictions around the country have shared data. My colleague, Julia Stoyanovich, as well as others in the community, has deep expertise in the burgeoning new area of fostering responsible data science, which fosters critical literacy, establishes values and beliefs as an explicit part of systems design and implementation, advances new technical research in methods, systems, and algorithms that reflect societal values and relevant laws, and reimagines new systems in data management and data sharing, which incorporates sociological questions about the people affected by algorithmic decision systems. The City Council could provide a list of approved secure environments within which agencies could share data.

**Second**, shared data is the necessary first step to better transparency, but it is not sufficient to effect change. Because most citizens encounter multiple city and state programs, it is necessary to link data across agency lines to fully understand the dynamic interaction of government with citizens. Agency capacity must be built to understand the many challenges associated with linking data cross agency lines, and to share knowledge about how best to address those challenges. Too few government employees have the requisite skills, and governments often do not have the salary flexibility to compete with the private sector to hire and retain enough in-house data analyst. Confidentiality rules often limit datasharing. Absent a clearly sufficient value proposition, it is difficult to obtain the resources necessary to surmount the many legal and technical hurdles that prevent cross-agency data collaborations. The time to reach agreement can take years—10 years in at least one case!

These combined challenges have led to the current Catch-22: because they cannot demonstrate the value of new data products, agencies cannot get the significant resources necessary to make use of linked data, but lack of resources mean that they cannot demonstrate value. In human terms, the cost of not combining data (in a timely fashion) is also evident. Dr. Leana Wen, Commissioner of Health, City of Baltimore has noted, as “part of Child Fatality Review, department heads in Baltimore City government get together once a month. We review every child death that happened in the city since the previous meeting. We ask what more we might have done to prevent that tragedy. In many cases, each of us has a file on the child or the family at least an inch thick. It’s



tragic to compare notes after the child has died—what more could we have done when the child was alive?”

Our experience is that training classes that bring agency staff together to solve such challenges in the context of addressing a common problem is critical to building such capacity. We have trained over 300 government agency staff from over 100 agencies across the country in executive education style classes. Three projects are highlighted on our website (<https://coleridgeinitiative.org/training>) and are available for download. The title of each of them show the nature of the work: “From Prosecuted to Job Recruited: An Exploratory and Machine Learning Approach to Employment after Prison”; “Addressing Recidivism: Intervening to Reduce Technical Violations and Improve Outcomes for Ex-Offenders” and “Mommy Don’t Go: Predicting and Preventing Recidivism of Mothers in the Illinois Criminal Justice System.” The City Council could enhance the capacity of agency staff to work with confidential data by encouraging agencies to send staff to similar executive education programs so that they learn the core data science skills necessary to do their jobs in the new data driven environment that we all face.

**Third**, much work has also been done to develop aggregate indicators that are understood by the agency and that are robust in construction, in health, housing, transportation and criminal justice, to name just a few. Agencies that are empowered to generate additional indicators, traceable to the underlying source data, will better understand the causes of, and how to respond to, changes in the indicators. Community feedback is, of course, essential. The aggregate data indicators could be disseminated to the community for comment and iterative feedback so that effective evidence-based policy can be developed around sensible aggregate indicators. The City Council could encourage agencies to be involved in such a feedback loop, and reward them for doing so.

In short, I recommend that City Council leverage substantial existing work that has been done at the federal and local level in three ways.

- (1) Ensure the right technology is put in place to protect confidentiality for data sharing.
- (2) Provide professional development funds adequate to ensure that agencies are engaged in the effort to change the way in which they do business.
- (3) Reward agencies that substantively engage with, and receive input from, the community in developing critical indicators.