

# Integrating Big Data and Thick Data to Transform Public Services Delivery

By Yuen Yuen Ang



Big data has revolutionized the delivery of business and government services. But what's missing? Government can greatly enhance the value of big data by combining it with "thick" data—rich qualitative information about users, such as their values, goals, and consumption behavior, obtained by observing or interacting with them in their daily lives. Having lots of big data can be overwhelming or have little utility if the data are "thin"—that is, they lack meaning for users or fail to capture issues that matter most. By yielding insights into what citizens really care about and how they consume services, thick data can inform both the collection and analysis of big data. Whereas big data is broad and thin, thick data is narrow and rich—blending them, therefore, yields a more holistic picture of the problem at hand.

In a new report from the IBM Center for The Business of Government, *Integrating Big Data and Thick Data to Transform Public Services Delivery*, the topic of "mixed analytics" is introduced. Mixed analytics can be defined as integrating big data and thick data to transform government decision making, public services delivery, and communication. This report presents three case

studies of organizations that employ mixed analytics at the international, federal, and city level, respectively.

These case studies include:

- The World Bank Social Observatory's p-tracking (participatory tracking) project among 32,000 village residents in India
- The APHIS (Animal and Plant Health Inspection Service, a division of the United States Department of Agriculture) tailored social marketing campaign
- The LA Express Park program, featuring dynamic pricing, in downtown Los Angeles

Together, this research offers a set of transferable lessons for agencies at all levels of government:

- Lesson 1: Big data is a means to an end, rather than an end.
- Lesson 2: Thick data can identify unexpected problems or previously unexpressed needs.
- Lesson 3: Thick data can inform the analysis of big data.
- Lesson 4: Mixed analytics can offer both scale and depth.
- Lesson 5: Applying technology is a social activity, not an isolated technical task.
- Lesson 6: The best solutions are not always high-tech.

The report concludes with five actionable recommendations for public managers:

1. Make data and technology relevant to the people who use it.
2. Leverage thick data at appropriate stages of the problem-solving process.
3. Build an interdisciplinary team of quantitative and qualitative experts who work closely with stakeholders.

*Yuen Yuen Ang is Associate Professor of Political Science at the University of Michigan, Ann Arbor. She is also a faculty associate at UM's Center for the Study of Complex Systems. In 2018, Ang was named an Andrew Carnegie Fellow by the Carnegie Corporation of New York.*

4. Combine big and thick data to improve communication.
5. Improve government agencies' knowledge of mixed research methods.

### Analytical Perspectives on Big data and Thick Data

Big data can be described as “massive data sets sifted by powerful analytical tools.”<sup>1</sup> Big data is defined not only by volume (the amount of data collected and analyzed), but also by velocity and variety. Velocity refers to the speed of data being produced, for example; streaming data is now widely available for real-time, instant analysis. Variety refers to the different forms of data collected, which can be

structured or unstructured, and the number of dimensions captured by the data.

If big data reflects volume, velocity, and variety for items that can be counted, thick data concerns information about the significance, meaning, and connections that humans assign to services or technologies, as well as the process by which they consume them. Thick data is generated through immersion in users' natural settings, rather than in laboratory-type settings.

Table 1 compares the attributes and functions of big data and thick data. Table 2 demonstrates the key characteristics of the three case studies examined.

**Table 1: Big data vs. Thick Data**

	Big Data	Thick Data
Format of data	Data in numerical format	Data in non-numerical, qualitative format
Volume	Usually large, even massive number of observations	Usually small number of observations
Methods of data collection	Digital records, digitized archival records, streaming data, transmission logs, numerical data scrapped from the Internet and social media	Participant observation, direct participant interviews, focus groups, small-n open-ended surveys, video recordings, qualitative data from the Internet and social media
Analysts	Research by social and computational scientists	Research by anthropologists and ethnographers
Immersion Requirement	Analysts need not be on-site to analyze the data	Usually on-site or direct observation online, and immersed in context (observing, participating, talking to people)
Role in problem solving	Generate solutions to largely known problems, such as by mapping behavioral patterns, making predictions, generating automated decisions	Identify problems that matter most to stakeholders, particularly unknown problems and previously unexpressed needs, and test solutions before scaling up
Strengths	Scale: generate insights that are generalizable to a large portion of or an entire population	Depth: identify what stakeholders care about in the first place; paint a holistic picture of certain experiences

Source: Yuen Yuen Ang

1. “The Power of Thick Data,” Wall Street Journal, March 21, 2014.

Table 2: Summary of key characteristics of case studies

	World Bank's Social Observatory	APHIS Public Outreach	LA Express Park
Actors and stakeholders	Interdisciplinary team of World Bank researchers, government actors at various levels in India (from leadership of Tamil Nadu down to villages), women's self-help groups	APHIS (agency) and researchers at Fors Marsh Group (contractor)	U.S. Department of Transportation (main funder), Los Angeles City, Xerox-PARC (contractor)
Target users	Rural residents, mostly women	Domestic and international travelers	Motorists
Big data source for quantitative analysis	Pilot census survey of 32,000 residents in Pudhu Vaazhvu, Tamil Nadu	AQIM (Agricultural Quarantine Inspection Monitoring) dataset	Real-time parking occupancy data from 6,000 smart meters
Thick data source for qualitative analysis	Researchers are embedded in communities to observe and listen to deliberations; direct participation by women self-help groups in designing survey	Input from agency experts at all stages of the research process; researchers visited airports and interviewed agency officers and travelers	Rapid ethnographic method (compressed periods of observation, interviews, direct participation, and videotaping)
Product	"P-tracking": survey designed with direct inputs from local women; collected data was visualized and shared with communities to inform deliberation and decision making	Targeted public outreach and social marketing campaign, with messages tailored by location and segment of travelers	Demand-based parking pricing, paired with mobile apps for viewing and booking parking spaces

Source: Yuen Yuen Ang



### Lessons Learned

- Big data is a means to an end, rather than an end.** The ongoing focus on big data may compel public managers to feel that they need to “do something” with the data, whether or not this is necessary or useful. In the U.S., recent federal plans that link big data to the nation’s “strategic assets” may unintentionally reinforce such pressures. But while public agencies should seriously consider big data as part of their toolkit, they should not use big data just for its own sake. All three case studies illustrate big data as a means to an end, rather than an end in and of itself.
- Thick data can identify unexpected problems or previously unexpressed needs.** Clearly, thick data and ethnography can complement big data analysis—the harder work arises in specifying how. Governments can miss obvious problems if they rely only on big data and analytics, as best illustrated by the case of LA Express Park. Thick data also proves particularly useful for informing the collection of big data, as seen in the World Bank’s Social Observatory project. If public agencies invest time and effort to collect big data without first inquiring what users care most about, they may measure the wrong things.
- Thick data can inform the analysis of big data.** Thick data also informs the analysis of big data. In the APHIS case, data scientists worked closely with agency experts to understand the construction and quality of the data sets, seeking inputs for the design of statistical models and interpretation of results. The data scientists who worked with APHIS found: “Because these results were to inform the development of a campaign that would not exist in a vacuum but instead would be implemented in a mix of policy, political, and budgetary influences, the final solution could not be determined by the machine-based analysis of the results alone
- Mixed analytics can offer both scale and depth.** Given the different advantages and functions of big data and thick data, the best research teams and technology designs typically use mixed analytics (big data and thick data) and mixed research methods (quantitative and qualitative). In addition, they feature an interdisciplinary team of specialists, not just data scientists. The World Bank brought in economists, sociologists, behavioral scientists, and information system specialists. Designing LA Express Park involved both engineers and ethnographers.

- **Applying technology is a social activity, not an isolated technical task.** Applying technology is often an intensely social activity, not a job that engineers and data scientists perform in isolation. At the World Bank, the Social Observatory team mobilized entire Indian villages at all stages of its program, from designing surveys, implementing them, to sharing data with the villagers. At APHIS, data scientists worked closely with agency experts, building rapport with them, incorporating their contextual knowledge into computational analysis, and even inviting officers to present the results to the agency's leaders to instill a sense of ownership.
- **The best solutions are not always high-tech.** Being “smart” does not necessarily entail using the most advanced technology available. Reiterating Lesson 1, public managers should focus on their goals or problems to be solved, and adopt a pragmatic approach that welcomes any mixture of solutions, whether high-tech or low-tech.
- **Combine big and thick data to improve communication.** The combination of big data and thick data is especially useful for improving communication, both public and internal. Ethnography can reveal which messages resonate most with citizens and why. And data scientists who converse with agency experts and observe them in their work environment can appreciate their challenges and goals. The most effective targeted marketing campaigns do not result from the biggest dataset—they come instead from data scientists who know their context and data well.
- **Improve government agencies' knowledge of mixed research methods.** Many governments around the world put a great deal of emphasis on the need for public servants to be “future ready” and data literate, but virtually none perceive the need for literacy in mixed research methods. Using big data well requires contextual knowledge; public managers must understand more than just numbers and analytics.



## Recommendations for Public Managers

- **Make data and technology relevant to the people who use it.** One common lesson that emerges from the case studies is not to use big data for its own sake. Instead, agencies should seek to make data and technology relevant to users, by combining technology with an “ask, observe, and immerse” approach.
- **Leverage thick data at appropriate stages of the problem-solving process.** Any problem-solving process has four steps: (1) Identify problems that matter to users or stakeholders, (2) propose solutions to the problems, (3) test and refine solutions on a small scale, and (4) implement solutions on a large scale. Qualitative research methods and thick data are especially useful for steps 1 and 3.
- **Build an interdisciplinary team of quantitative and qualitative experts who work closely with stakeholders.** Big data initiatives should not involve only data scientists, but should also have input from qualitative researchers or ethnographers. Ideally, such an interdisciplinary team should divide their work along the lines of recommendation 2, and it should work closely with stakeholders and clients.

**Integrating Big Data and Thick Data to Transform Public Services Delivery**  
by Yuen Yuen Ang



[businessofgovernment.org](http://businessofgovernment.org)



[businessofgovernment@us.ibm.com](mailto:businessofgovernment@us.ibm.com)



202) 551-9342

