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.” 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](image)

Source: Yuen Yuen Ang

**管理**

### 课程：案例研究的关键特征

<table>
<thead>
<tr>
<th>年度银行的社会观察室</th>
<th>APHIS 公共宣传</th>
<th>LA 表达公园</th>
</tr>
</thead>
<tbody>
<tr>
<td>行动者和利害关系人</td>
<td>多学科研究小组，年年度银行研究人员，各级政府官员 (包括泰米尔纳德邦领导的自决到村庄），女性自助组织</td>
<td>APHIS (农业) 和研究人员 Fors Marsh Group (承包商)</td>
</tr>
<tr>
<td>目标用户</td>
<td>农村居民，主要是女性</td>
<td>国内和国际旅行者</td>
</tr>
<tr>
<td>大数据分析</td>
<td>模拟人口普查调查 32,000 人</td>
<td>AQIM (Agricultural Quarantine Inspection Monitoring) 数据集</td>
</tr>
<tr>
<td>充实数据分析</td>
<td>研究人员嵌入社区观察和参与；直接参与女性自助组织在设计调查</td>
<td>从各个阶段获取机构专家的输入；研究人员走访机场并采访机构官员和旅行者</td>
</tr>
<tr>
<td>产品</td>
<td>“P-tracking”：调查设计直接来自当地女性；收集的数据可视化并分享给社区以进行决策制定</td>
<td>针对特定地点和旅行者群体的公共宣传和社会营销运动，结合移动应用程序用于查看和预订停车位</td>
</tr>
</tbody>
</table>

来源：Yuen Yuen Ang

### 学习到的教训

- **大数据是手段，而不是目标。** 这样的持续聚焦可能迫使管理者感觉“需要做点什么”与数据有关，无论是必要的还是有用的。在美国，最近联邦计划将大数据与国家的“战略资产”联系起来可能无意中加剧了这样的压力。但同时，政府机构应认真考虑大数据作为其工具库的一部分，而不能仅仅为了大数据而大数据。所有的案例研究都说明了大数据作为手段而非目标的性质。

- **充实数据可以帮助识别意外问题或未表达需求。** 显然，充实数据和民族志可以补充大数据分析——即更难的工作体现在具体化上。政府可能会忽略明显的问题如果它们仅仅依赖于大数据和分析，如LA 表达公园的案例所示。充实数据同样有助于了解数据的收集，如年度银行的社会观察室项目。如果公共机构投入时间和努力收集大数据而没有首先询问什么是用户最关心的，则他们可能会做错事情。

- **充实数据可以促进大数据的分析。** 大数据也帮助分析大数据。在APHIS案例中，数据科学家紧密与部门专家合作以理解数据集的构造和质量，寻求针对设计和结果解释的输入。数据科学家与APHIS合作时发现：“因为这些结果将用于创建一个不会孤立存在的运动，而将被整合到政策、政治和预算影响中，最终的解决方案不能仅由机器分析的结果确定。”

- **混合分析可以提供规模和深度。** 给定大数据和充实数据的不同优势和功能，最好的研究团队和科技设计通常使用混合分析（大数据和充实数据）和混合研究方法（定量和定性）。此外，它们包括跨学科团队专家，而不仅仅是数据科学家。世界银行带来了经济学家、社会学家、行为科学家和信息系统专家。设计LA 表达公园涉及工程师和民族志研究人员。
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.

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.

• 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.

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

businessofgovernment.org
businessofgovernment@us.ibm.com
202) 551-9342