Promoting Data Access and Data Protection

How emerging technologies can help agencies navigate the triple constraint of data.

Michael Hawes
Senior Advisor for Data Access and Privacy
U.S. Census Bureau

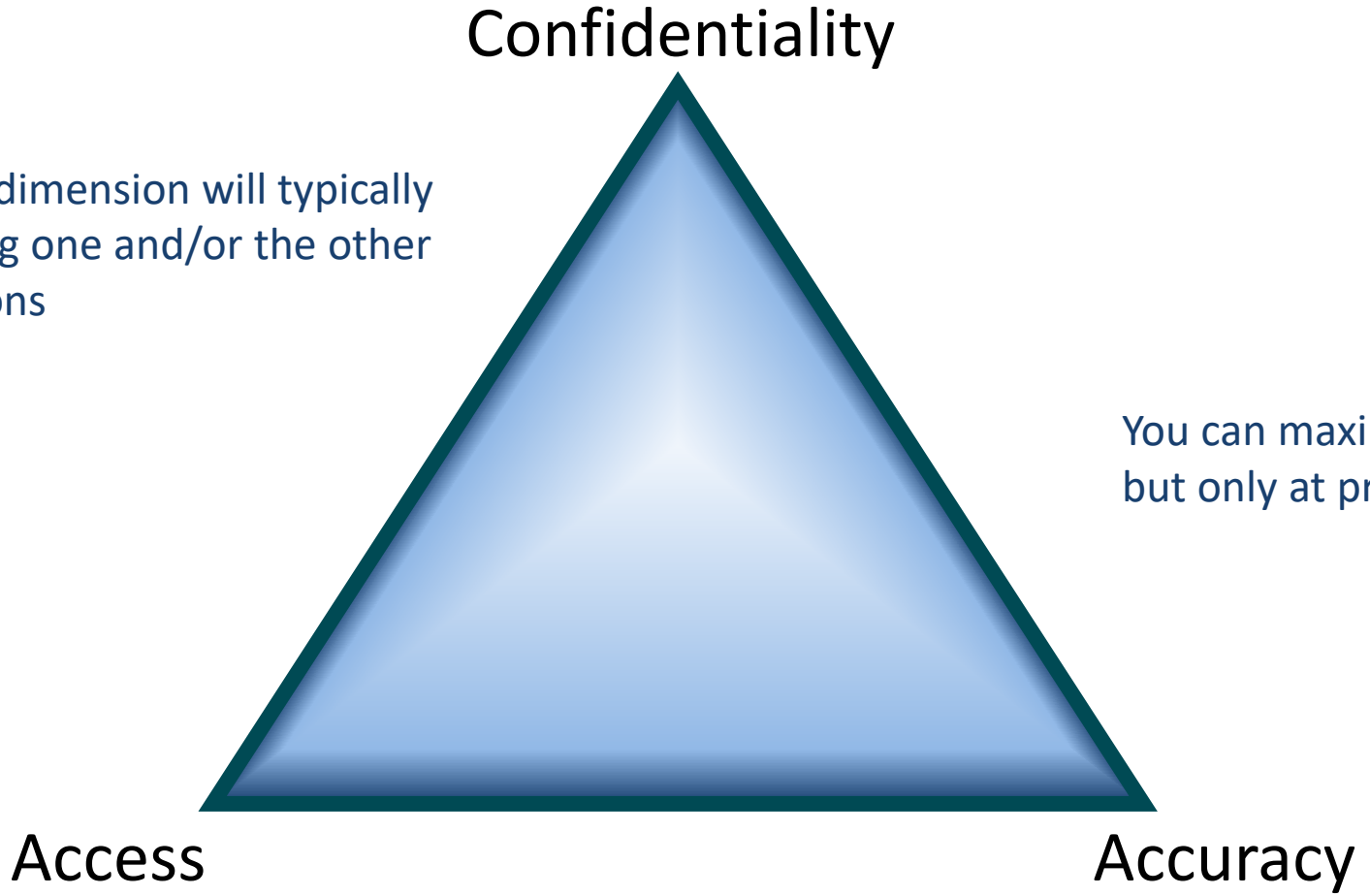
Every time you release any statistic calculated from a confidential data source you "leak" a small amount of private information.

If you release too many statistics, too accurately, you will eventually reveal the entire underlying confidential data source.

Dinur, Irit and Kobbi Nissim (2003) "Revealing Information while Preserving Privacy" PODS, June 9-12, 2003, San Diego, CA

The Triple Constraint of Data

Increasing one dimension will typically require reducing one and/or the other of the dimensions



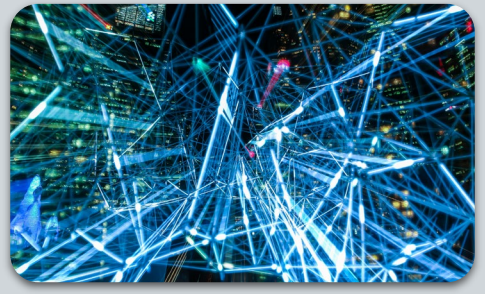
You can maximize on two dimensions, but only at profound cost to the third.

Data Access and Disclosure Risk

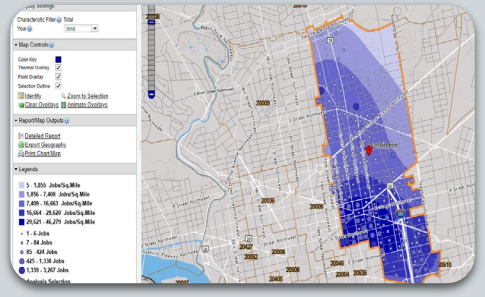
Supporting increased access to agency data for evidence-building increases disclosure risk in a number of ways:

- **Increased quantities of data being released;**
- **Increased demand for microdata;**
- **Increased demand for highly granular data; and**
- **Increased demand for direct identifiers (to permit linkage).**

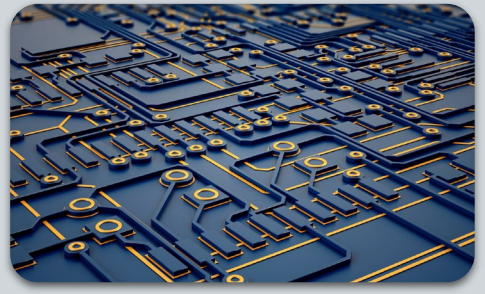
Emerging Technologies Can Help



Synthetic Data



Query Systems



Secure Multiparty Computation

Differential Privacy

But…

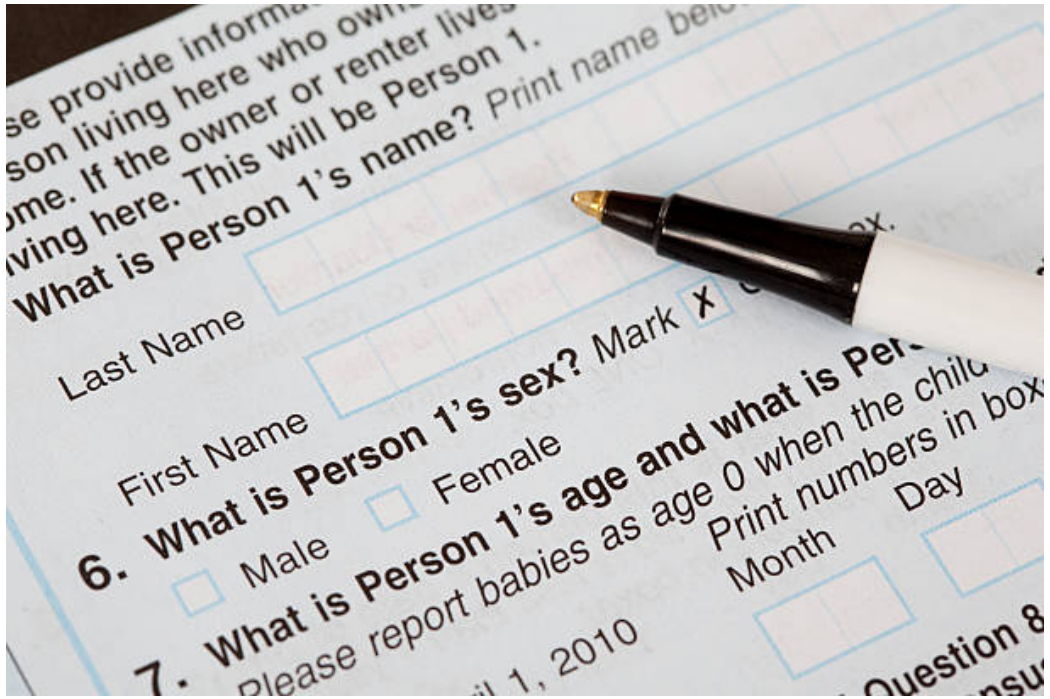
None of these emerging technologies is (currently) an "off the shelf" solution.

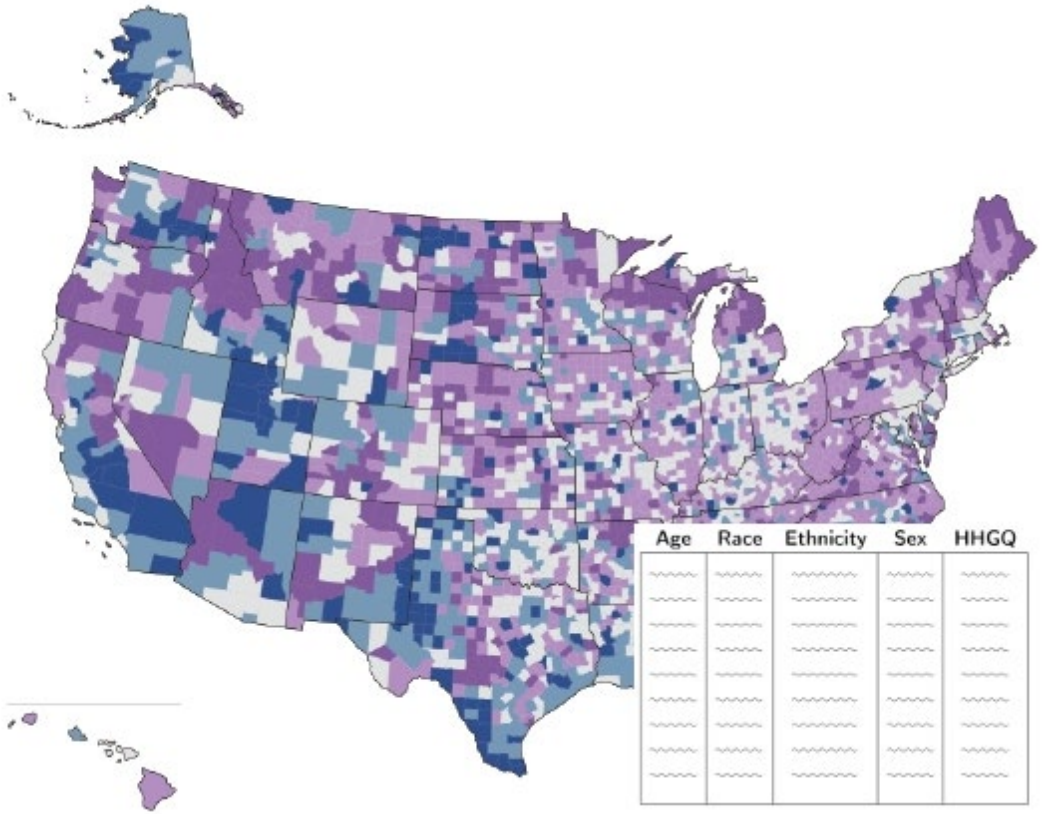
Proper implementation requires:

- **Statistical and technical expertise;**
- **Engagement with data users;**
- **Configuration and optimization; and**
- **Up front agency investment.**

Plus, many of these technologies still require additional research and development to have broad applicability.

The U.S. Census Bureau's Experience with Differential Privacy





Michael Hawes
Senior Advisor for Data Access and Privacy
U.S. Census Bureau

301-763-1960 (Office)
michael.b.hawes@census.gov