Towards a Principled Disclosure Avoidance Framework:

Identifying the Characteristics of an Ideal, Applied Disclosure Avoidance System

Joint Statistical Meetings
August 5, 2024

Michael B. Hawes
Senior Statistician for Scientific Communication
U.S. Census Bureau

The views expressed in this presentation are those of the presenter and not the Census Bureau.

United States® Census Bureau

Towards a Principled Framework for Disclosure Avoidance

Michael B Hawes[1], Evan M Brassell[1], Anthony Caruso[1], Ryan Cumings-Menon[1], Jason Devine[1], Cassandra Dorius[1,2], David Evans[1,3], Kenneth Haase[1], Michele C Hedrick[1], Scott H Holan[1,4], Cynthia D Hollingsworth[1] Eric B Jensen[1], Dan Kifer[1,5], Alexandra Krause[1], Philip Leclerc[1], James Livsey[1], Roberto Ramirez[1], Rolando A Rodríguez[1], Luke T Rogers[1], Matthew Spence[1], Victoria Velkoff[1], Michael Walsh[1], James Whitehorne[1], and Sallie Ann Keller[1,3]

[1]U.S. Census Bureau*, [2]Iowa State University, [3]University of Virginia, [4]University of Missouri, [5]Penn State University

Draft, June 28, 2024

Manuscript under review by the Harvard Data Science Review

An earlier version of this manuscript was presented at the May 2024 NBER Workshop on Data Privacy Protection and the Conduct of Applied Research

Any opinions or viewpoints are the authors' own and do not reflect the opinions or viewpoints of the U.S. Census Bureau.

United States® Census Bureau

The Triple Tradeoff of Official Statistics

The more statistics you publish, and the greater the granularity and accuracy of those statistics, the greater the disclosure risk.

All statistical techniques to protect confidentiality impose a tradeoff between the degree of data protection and the resulting availability and accuracy of the statistics.

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

Disclosure Avoidance Techniques and the Triple Tradeoff

Nearly any disclosure avoidance method can be applied to implement very different balances along the three dimensions, depending on the implemented parameters selected.

Suppression

- *Cell size thresholds, p% rules*

Coarsening

- *Rounding rules (e.g., 3, 10, 1000)*

Swapping

- *Swap keys, rates, geographies*

Differential Privacy

- *Privacy-loss budgets and allocations*

Objective

We need a set of overarching principles that an ideal, applied disclosure avoidance system should meet…

…while distinguishing those principles from any choices relating to the implementation of that system.

What is a Disclosure Avoidance System?

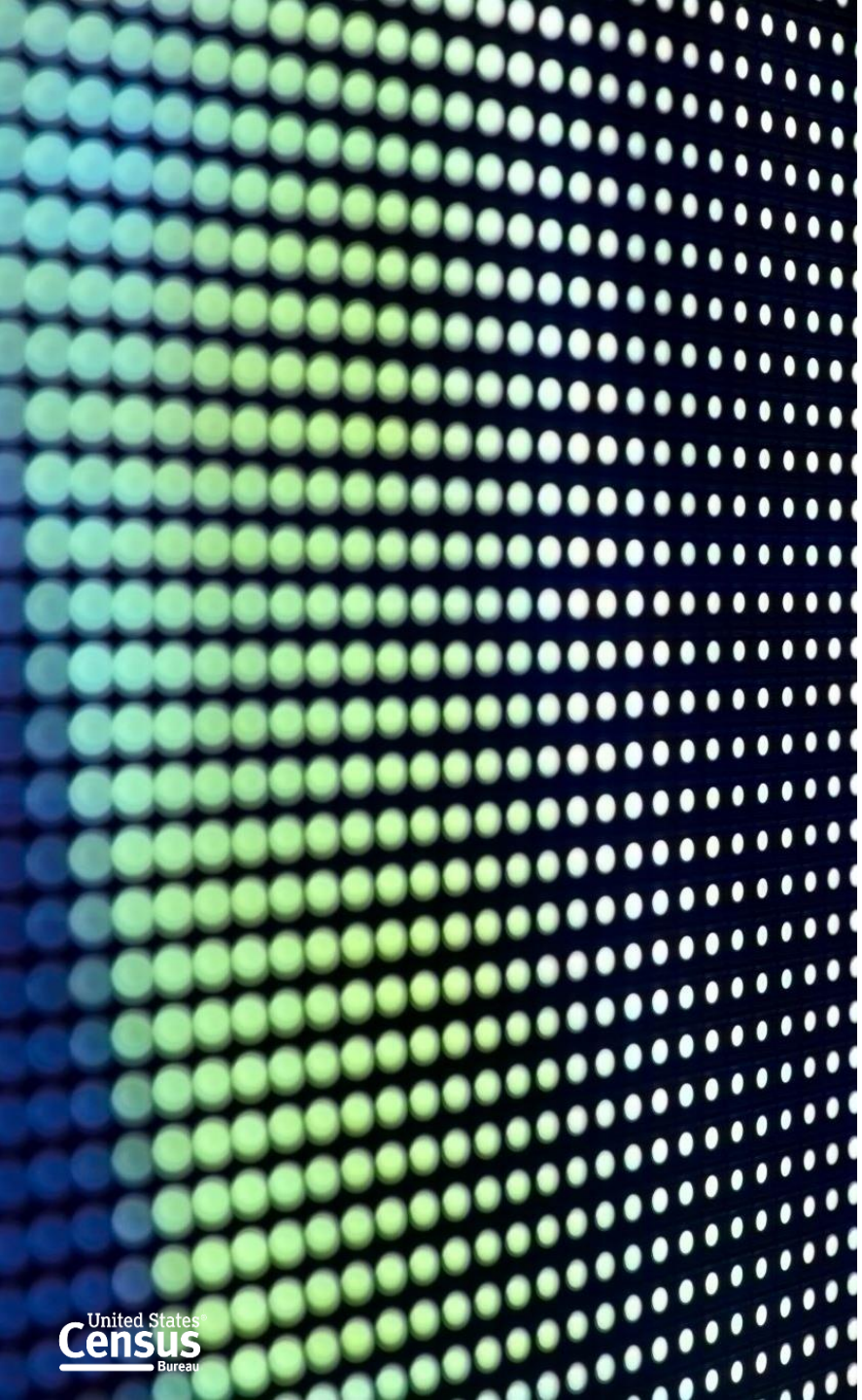
A Disclosure Avoidance System is a set of one or more statistical methods that transform confidential information (or data derived from confidential information) from or about individual data subjects into statistics that describe, estimate, or analyze the characteristics of groups, without identifying the data subjects that comprise such groups.

What is an Applied Disclosure Avoidance System?

An applied Disclosure Avoidance System is one that performs within the operational realities and production cycles of a national statistical office. As such, it acknowledges and is adaptable to requirements stemming from the legal, policy, scientific, resource, and stakeholder environments within which it is operating.

What is an Ideal, Applied, Disclosure Avoidance System?

An ideal, applied, Disclosure Avoidance System is one that conforms to a set of overarching principles or features relating to the efficiency, effectiveness, and flexibility of the system as it transforms confidential information from (or about) data subjects into quality statistics for public release.



Distinguishing these Principles from Implementation Choices

Principles

Reflect characteristics that all DA systems should ideally have, regardless of the specific technology or disclosure limitation mechanism being employed.

Should be universally applicable regardless of the type, format, or context of data being protected.

Implementation

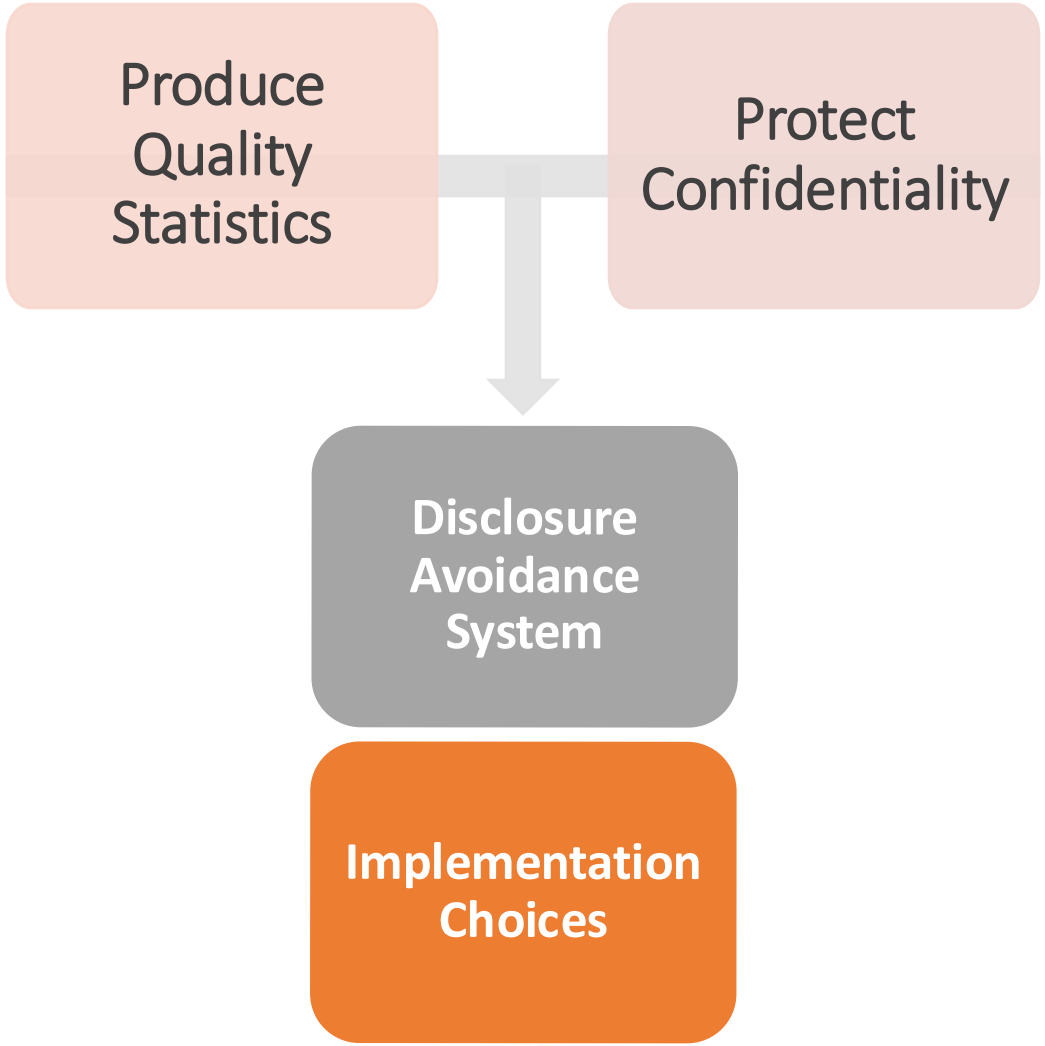
Reflect specific choices about the appropriate balance between data privacy, accuracy, and availability.

Should be informed by the characteristics of the data, the context of the statistical product, and the intended objectives and requirements of the agency and its stakeholders.

Balancing Our Legal Obligation with Our Primary Mission

The Census Bureau's mission is to produce high quality statistics and statistical products.

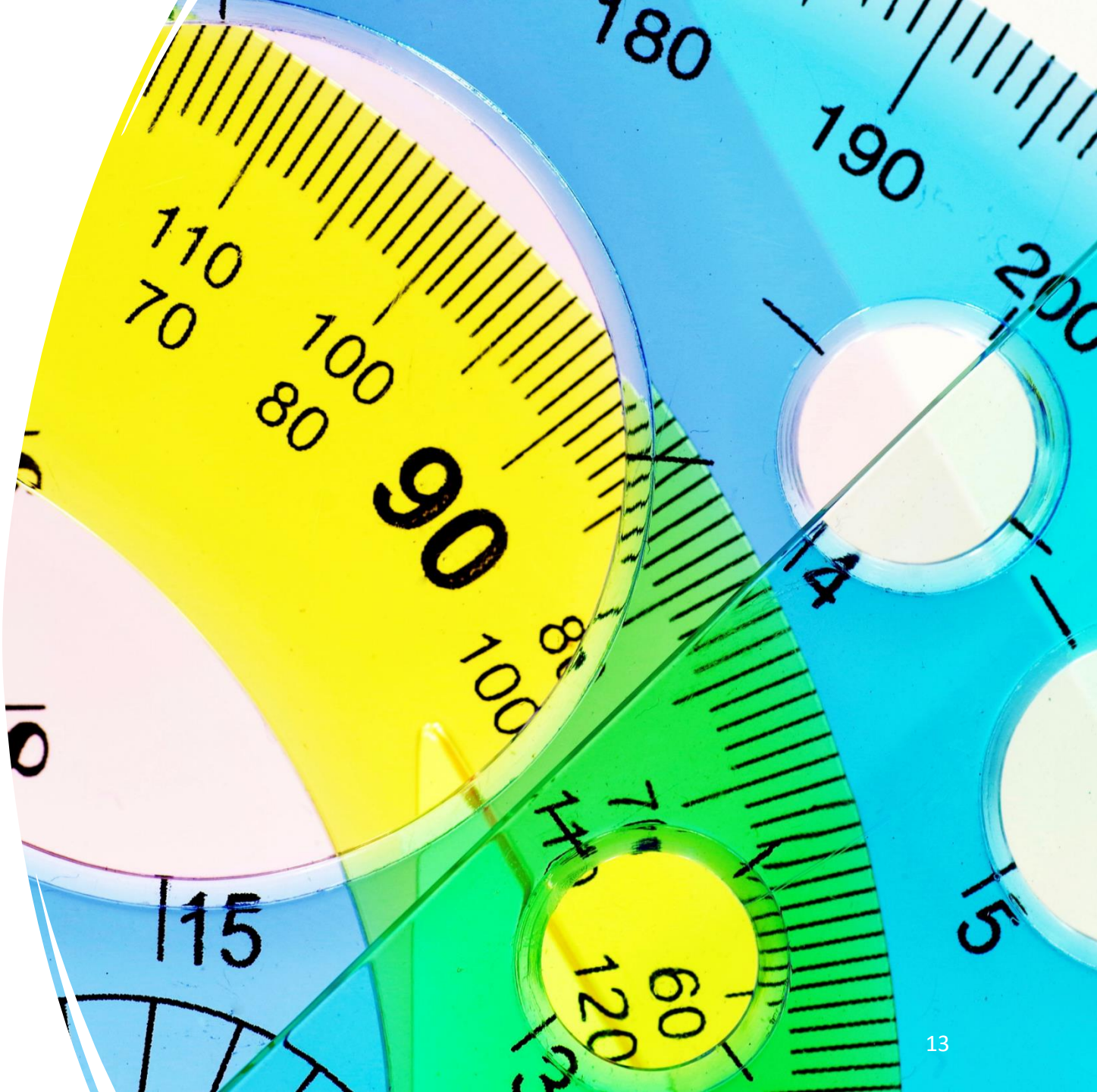
Title 13's confidentiality protections were established in support of that mission.



Characteristics of an Ideal, Applied Disclosure Avoidance System

Principle #1

It should support meaningful assessment of disclosure risk



Principle #2

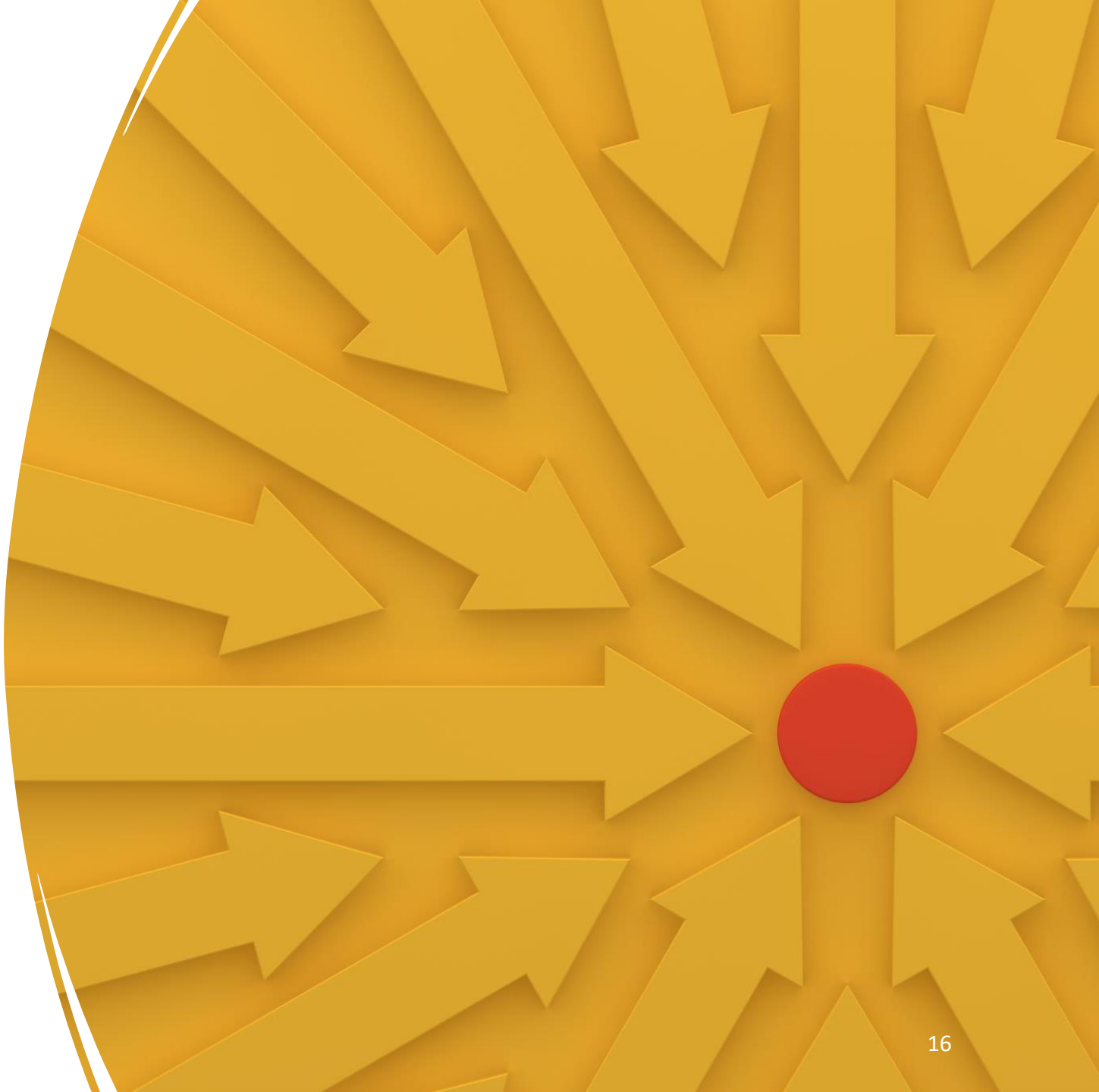
It should support meaningful assessment of the impact of data protections on the quality of statistics

Principle #3

It should be able to target protection

Principle #4

It should be able to target data quality



Principle #5

It should track cumulative disclosure risk over time

Principle #6

It should be transparent

Principle #7

It should be feasible

The 7 Principles of an Ideal, Applied Disclosure Avoidance System

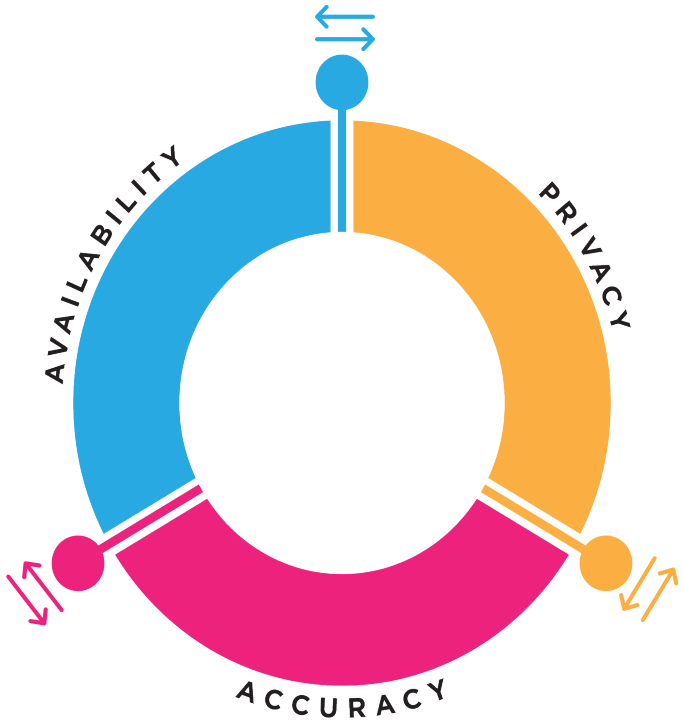
1. Support meaningful assessment of disclosure risk
2. Support meaningful assessment of the impact of data protections on the quality of statistics
3. Target protection
4. Target data quality
5. Track cumulative disclosure risk over time
6. Be transparent
7. Be feasible

There can be tension between these principles

1. Support meaningful assessment of disclosure risk
2. Support meaningful assessment of the impact of data protections on the quality of statistics
3. Target protection
4. Target data quality
5. Track cumulative disclosure risk over time
6. Be transparent
7. Be feasible

Examples of implementation choices independent of the selection of a DAS

- Desired balance of accuracy, confidentiality, and availability of statistics
- Targeting and prioritizing of confidentiality protections
- If any data elements should be excluded from protection
- Prioritization of use cases
- Statistical product design requirements
- Operational considerations



Questions and Discussion

United States Census Bureau