



Towards a Principled Discussion of a Disclosure Avoidance Framework

Identifying the Characteristics of an Ideal, Applied Disclosure Avoidance System

Michael B. Hawes

Senior Survey Statistician for Scientific Communication
U.S. Census Bureau

NBER Conference on Data Privacy Protection and the Conduct of
Applied Research: Methods, Approaches and their Consequences
05/16/2024

Towards a Principled Discussion of a Disclosure Avoidance Framework: Identifying the Characteristics of an Ideal, Applied Disclosure Avoidance System

Michael B Hawes¹, Evan M Brassell¹, Anthony Caruso¹, Ryan Cumings-Menon¹, Jason Devine¹, Cassandra Dorius^{1,2}, David Evans^{1,3}, Kenneth Haase¹, Michele C Hedrick¹, Scott H Holan^{1,4}, Cynthia D Hollingsworth¹, Eric B Jensen¹, Dan Kifer^{1,5}, Alexandra Krause¹, Philip Leclerc¹, James Livsey¹, Roberto Ramirez¹, Rolando A Rodríguez¹, Luke T Rogers¹, Matthew Spence¹, Victoria Velkoff¹, Michael Walsh¹, James Whitehorne¹, and Sallie Ann Keller^{1,3}

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

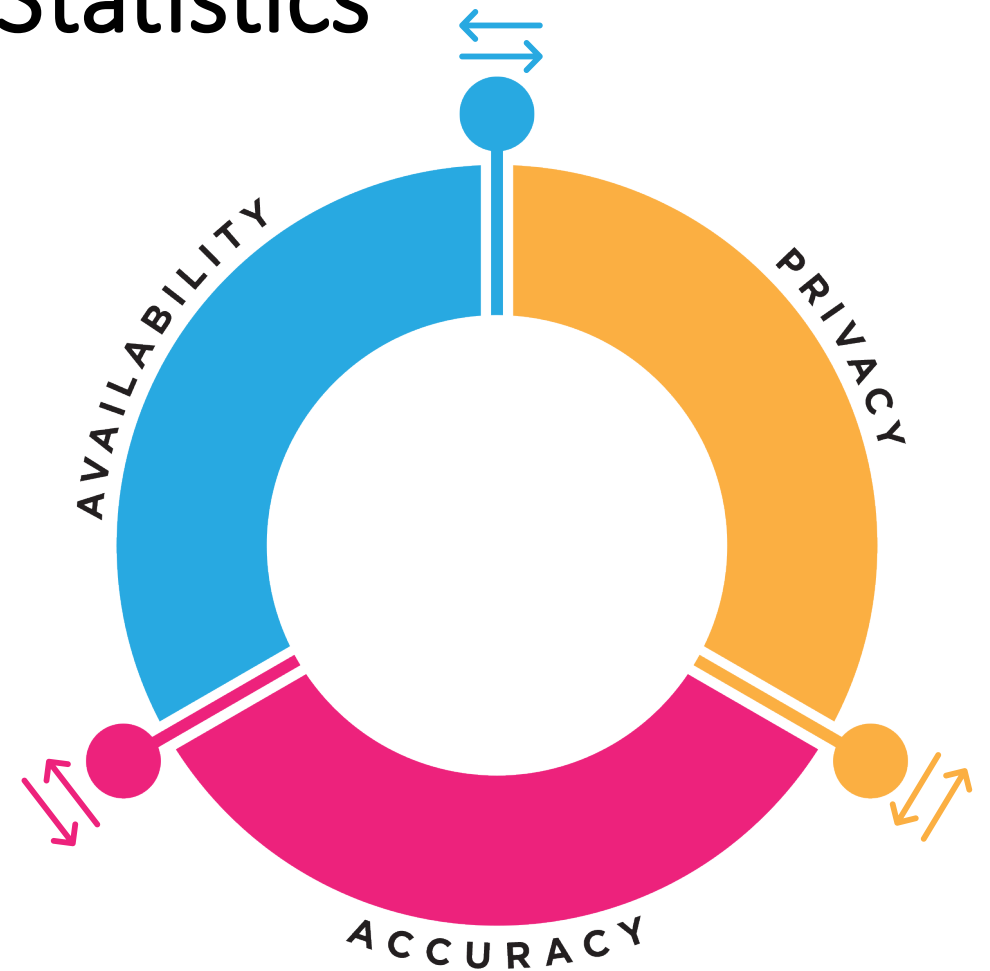
Draft, May 14, 2024

The Triple Trade-Off 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

The selection of a particular disclosure avoidance (DA) technique does not directly impact agency decision-making within the context of the triple tradeoff. Nearly any DA technique can be applied to implement very different balances along these three dimensions, depending on the implemented parameters selected.

Example DA Techniques	Examples of Parameters that Implement the Triple-Tradeoff 
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.

Disclosure Avoidance Systems accomplish this through the application of statistical disclosure limitation techniques to reduce (but not eliminate) disclosure risk in the statistical products being produced.

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

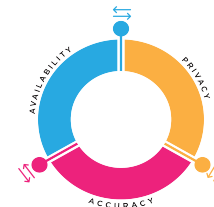
Principles

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.

Reflect specific choices about the appropriate balance between data availability, utility, and confidentiality.

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.



Implementation



IMPORTANT

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.

We must remember that disclosure avoidance is a legal obligation and a necessary activity, but it is one that we undertake in tandem with, and in support of, our primary mission to produce quality statistics.

Produce Quality Statistics

Protect Confidentiality

Disclosure
Avoidance System

Implementation
Choices

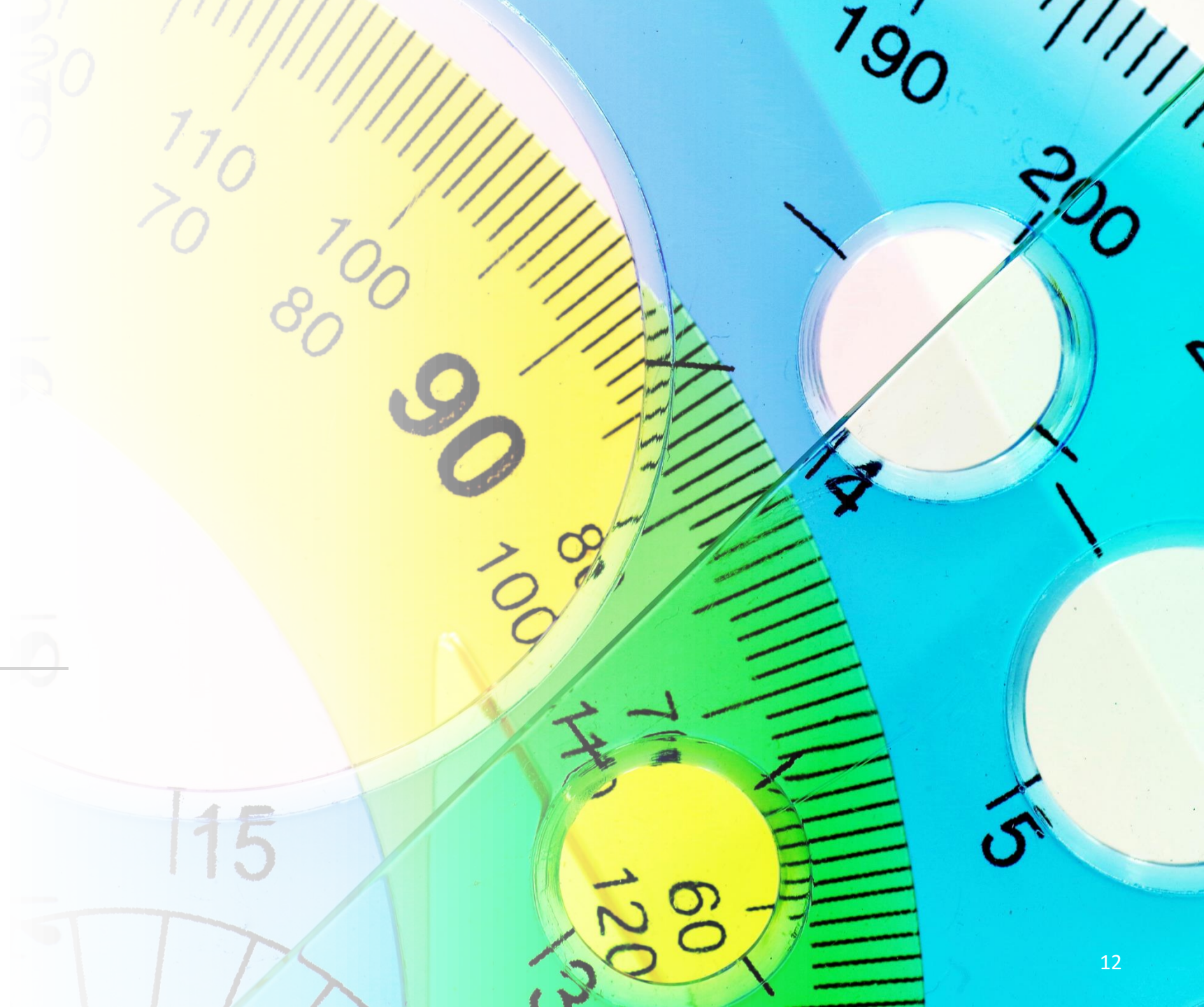


Characteristics of an Ideal, Applied Disclosure Avoidance System



Principle #1

It should be able to
assess disclosure risk



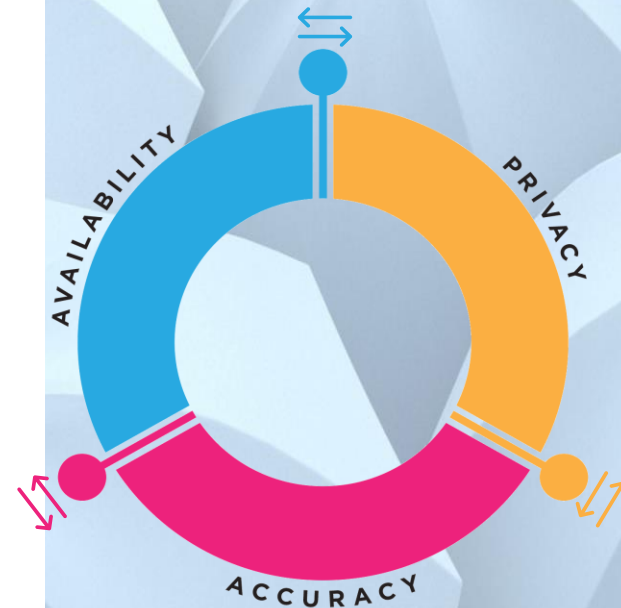
Principle #2

It should be able to assess the impact of data protections on the quality of statistics



Principle #3

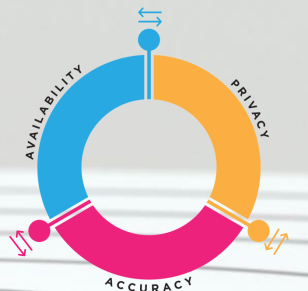
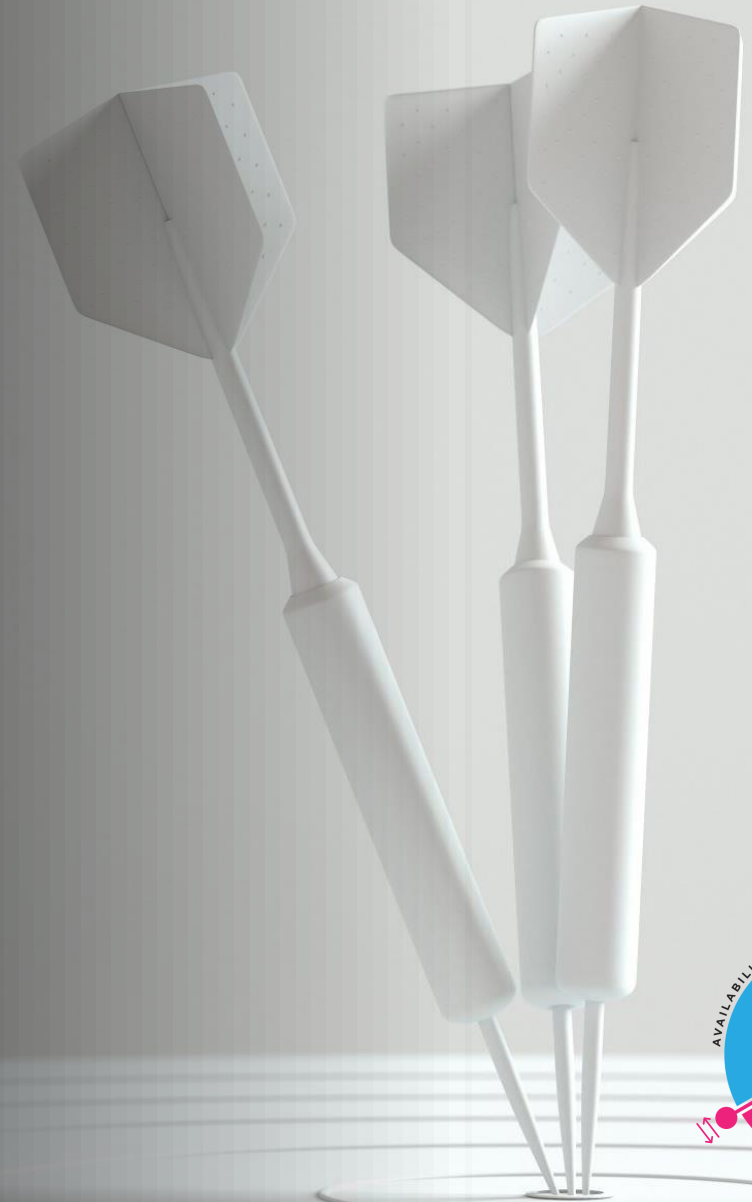
It should be able to target protections for vulnerable data subjects





Principle #4

It should be able to achieve quality targets for statistical products



Principle #5

It should track cumulative disclosure risk over time



Principle #6

It should be transparent



Principle #7

It should be feasible



Distinguishing these Principles from Implementation Choices

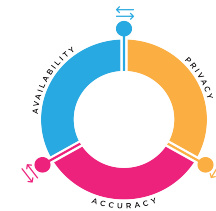
Principles

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.

Reflect specific choices about the appropriate balance between data availability, utility, and confidentiality.

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.



Implementation

Examples of implementation choices independent of the selection of a DAS

Desired balance of accuracy, confidentiality, and availability of statistics

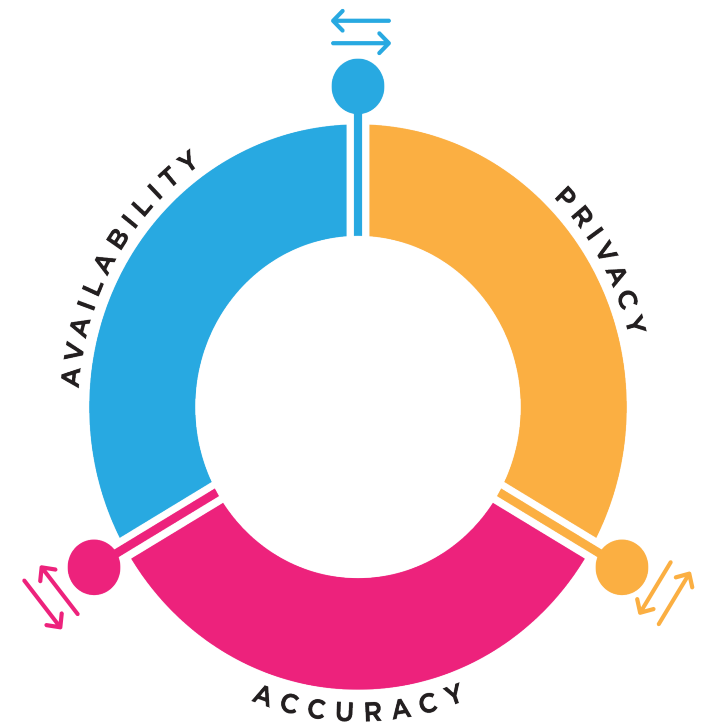
Targeting and prioritizing of confidentiality protections

If any data elements should be excluded from protection

Prioritization of use cases

Statistical product design requirements

Operational considerations





The ideal DAS should:

- assess disclosure risk;
- assess impact on quality of statistical products;
- be able to target protections for vulnerable data subjects;
- be able to achieve quality targets for statistical products;
- track cumulative disclosure risk over time;
- be transparent;
- be feasible.

Using the Principles

- There can be tension between these principles

Questions and Discussion

