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

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, May 14, 2024

1 Introduction

The principal function of a statistical agency is to produce and release quality statistics. In pursuing that mission, most statistical agencies have a countervailing obligation to protect the confidentiality of the data entrusted to them by their data subjects. While an agency can leverage many different statistical techniques and algorithmic frameworks to provide confidentiality protection, it is not always clear which approach is best for protecting a particular statistical product while maintaining the agency's broader objectives for disseminating the product. Without a clear set of criteria to use when comparing systems, selecting one approach over another can often be challenging and occasionally controversial. This paper seeks to establish a set of objective principles that can inform statistical agencies' evaluation and selection of disclosure avoidance systems.

2 What is an Ideal, Applied Disclosure Avoidance System?

A disclosure avoidance system (DAS) 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 the characteristics of groups without identifying the data subjects that comprise such groups. Disclosure avoidance systems accomplish this by applying statistical disclosure limitation (SDL) techniques to reduce (but not eliminate) disclosure risk in the statistical products being produced.

An *applied* DAS is one that performs within the operational realities and production cycles of a national statistical office. As such, it acknowledges and adapts to requirements stemming from the legal, policy, scientific, resource, and stakeholder environments it operates within. This can mean being able to adapt to and effectively handle exogenously determined requirements such as data product design and schedule constraints.

An *ideal*, applied DAS is one that both adapts to external requirements and 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.

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

3 The Triple Trade-off of Official Statistics

At the heart of the challenge relating to disclosure avoidance is a set of interrelated constraints known as the "Triple Trade-off of Official Statistics" (Abowd and Hawes, 2023a,b; Hawes, Michael B., 2021). Every statistic an agency publishes that is derived from a confidential data source will reveal or leak confidential information in the process(Dinur and Nissim, 2003). Consequently, the more statistics an agency publishes, and the greater the granularity and accuracy of those statistics, the higher the risk of disclosure. Conversely, every statistical technique used to protect confidentiality will degrade the resulting statistics in some manner. This is not a side effect: it is how SDL methods protect confidentiality. As a result, statistical agencies have to navigate an inherent trade-off between the degree of data protection and the resulting availability and accuracy of the statistics to be released. When navigating this trade-off, agencies must ask themselves questions such as "How much disclosure risk is too much?" or "How accurate, and at what level of granularity, do these statistics need to be to meet their intended (or legally mandated) uses?"

4 Inherent Features versus Parameter Choices

When evaluating the strengths and limitations of different disclosure avoidance approaches, it is essential to distinguish the characteristics of the systems from the decisions and choices made under the Triple Trade-off that will be implemented through those systems. For example, a disclosure avoidance system based on complementary cell suppression can offer a high degree of data protection (with low data availability and quality) if it is implemented with a high primary suppression threshold (e.g., suppress values under 10,000), or it can offer a low degree of protection (with much higher data availability and quality) if it is implemented with a low primary suppression threshold (e.g., suppress values less than or equal to 3). Similarly, a rounding-based system can provide higher accuracy with less robust protections or lower accuracy with stronger protections, depending on whether it rounds values to the nearest 10 or 10,000. In this context, selecting a disclosure avoidance system based on suppression over one based on rounding should hinge on the characteristics and the relevant pros and cons inherent to those particular statistical disclosure limitation methods rather than on specific implementation or parameter choices for those systems under the Triple Trade-off. Put another way, agency decision-makers can have a legitimate debate about the relative merits of suppression-based versus rounding-based disclosure avoidance systems, but it would be unfair for them to evaluate a suppression-based system with a primary suppression threshold of 10,000 against a rounding-based system that rounds to the nearest 10.

The selection of both a DAS and parameters representing the agency's chosen balance under the Triple Trade-off are crucial decisions with tangible consequences for the resulting value of the statistical product to be released. An informed dialogue and debate about these choices should meaningfully distinguish between the system's characteristics and parameter selections.

5 Characteristics of an Ideal, Applied Disclosure Avoidance System

5.1 (Principle 1) It Should Assess Disclosure Risk

Central to the ability of a DAS to function is the ability to assess disclosure risk. After all, how can an agency protect its data against unauthorized disclosure if it cannot determine how likely a disclosure is to occur?

In this context, however, the agency must be clear about what constitutes a disclosure and what does not. Many federal confidentiality statutes approach disclosure from the perspective of revealing information contained in a data subject's census or survey response (or in the context of administrative records or other third-party sources, information maintained by the statistical agency in the data subject's records). Section 9(a) of Title 13, for example, prohibits any data release "whereby the data furnished by any particular establishment or individual under this title can be identified." Similarly, Section 3563 of the Foundations for Evidence-Based Policymaking Act of 2018 requires recognized statistical agencies and units to "protect the trust of information providers by ensuring the confidentiality and exclusive statistical use of their responses."

A simple reading of these requirements would seem to indicate that agencies would be prohibited from releasing any statistical products that would allow an attacker (someone trying to undermine an agency's disclosure avoidance mechanisms to learn about a data subject) to determine a data subject's exact response to any particular census or survey question with complete certainty.

Given the statutory frameworks, revealing a particular data subject's actual census or survey response would be an indisputable disclosure. Likewise, releasing a statistical product that permits deduction or inference about a data subject's response with perfect certainty would also be a clear disclosure. What about inference about a data subject's response with less than perfect certainty? If an agency merely modified one attribute on one data subject's record and then published the resulting data with the caveat that "one or more records have been modified," would that trivial amount of uncertainty be sufficient to "protect" confidentiality under these statutes? If 99.99999% certainty would constitute a disclosure, what about 99% or 95%? What about a coin toss (roughly 50%)? Here, we venture away from the textual letter of the law and wander into the realm of agency interpretation of their statutory responsibilities. In doing so, we have also ventured away from the responsibilities of an ideal DAS and into the territory of decision-making within the confines of the Triple Trade-off.

It is also important to note that not all inferences are equally concerning. Statistical agencies exist to produce statistical products that inform inferences about their respective societies. As such, the principal purpose of agencies' statistical products is to permit inference about those societies and their various groups and communities. Because neither membership in those groups nor the statistical characteristics of those groups are purely random, legitimate inference about those groups can inherently improve inference about a specific data subject's census or survey response. In the 2020 Census, for example, the racial and ethnic makeup of the state of Maine was 90% non-Hispanic White. This simple statistic about the demographic environment of Maine would allow an attacker to infer with 90% certainty that a randomly-selected data subject within the state is non-Hispanic White. In assessing the impact of inference on disclosure risk, an ideal DAS must be able to differentiate inferences based on broader societal information that does not rely on the particular data subject's response from inferences informed by leakage from the data subject's record. Only the latter type of privacy-eroding inferences should be considered a disclosure at whatever exogenously-defined (i.e., agency-determined) level of certainty.

Having clarified what disclosure is (and is not) within the context of disclosure avoidance systems, it is now vital for a disclosure avoidance system to be able to assess the degree to which the uncertainty introduced by the DAS can mitigate that underlying disclosure risk. Otherwise, without the ability to assess disclosure risk and to measure uncertainty, there would be no way of knowing how much uncertainty would need to be introduced to achieve a desired level of risk mitigation. It should be noted, however, that the ability to assess disclosure risk and to measure uncertainty—both essential features of this principle—in no way prescribe or dictate how much risk mitigation is necessary or desirable. Put another way, the ability to assess disclosure risk and the corresponding level of uncertainty from SDL is a prerequisite for agencies to navigate the privacy/accuracy/availability dimensions of the Triple Trade-off, but this feature of an ideal DAS does not determine the optimal point along any of those dimensions.

The underlying disclosure risk of a statistical product protected by a DAS can be assessed in different ways. Some systems can provide assessments of disclosure risk through analytical measures of the underlying uncertainty inherent to a statistical product's design and the additional uncertainty afforded by the application of SDL. Some coarsening SDL techniques, like uncontrolled, interior-cell rounding routines, can ensure that the most granular data disaggregations, from which all higher aggregations are built, limit precision to a specified degree of uncertainty. Similarly, properly implemented complementary suppression routines can ensure that no suppressed cell can be recalculated within a given degree of precision. Formally private SDL approaches can also enable analytic assessments of uncertainty by measuring the underlying leakage of response information in each released statistic. In each of these examples, the ability to assess uncertainty is independent of the agency's view on how much uncertainty is necessary or desirable. Rounding parameters, cell suppression thresholds, or formal privacy-loss budgets can be set high, low, or anywhere between. The challenge with these analytical approaches to uncertainty assessment, however, is that the uncertainty they measure does not necessarily directly correlate with tangible disclosure risk. Increasing uncertainty diminishes disclosure risk, but the ability of attackers to draw accurate inferences about data subjects in the presence of different levels and types of uncertainty also depends on the technology and external information available to the attacker. Some mathematically possible attacks may not represent credible threats to an

agency's disclosure avoidance protections because their underlying assumptions about computing power and external data may be unrealistic within reasonable time horizons.

An alternative method of assessing disclosure risk is through empirical assessment by simulating an actual attack on the statistical product's confidentiality protections and attempting to reveal or re-identify information about particular data subjects accurately. The most common form of these empirical assessments is re-identification studies on public-use microdata files or reconstruction-abetted re-identification studies on tabular products. However, increasingly sophisticated forms of simulated attacks, like membership inference attacks or likelihood-ratio test inference attacks, are also valuable techniques. The challenge with these empirical assessments is that because they are modeled on specific attack vectors, they are inherently limited by the assumptions an agency makes about the assessed attack vector. Achieving low re-identification rates through one of these assessments does not mean that the underlying disclosure risk of a statistical product is low in reality–it merely means that there is a low disclosure risk for the product against a particular attack vector under the prespecified assumptions and at the present moment.

The current limitations of both empirical and analytical approaches to assessing disclosure risk mean that no existing disclosure avoidance system can perfectly achieve this principle. Despite these limitations, however, some systems will be inherently better than others at achieving this principle's objectives. Properly applied uncontrolled rounding mechanisms can better ensure baseline degrees of uncertainty than controlled rounding mechanisms. Linear programming-based complementary suppression can provide guarantees that simple primary suppression algorithms cannot. Moreover, formally private noise-injection can provide provable measures of uncertainty that other forms of noise injection cannot.

5.2 (Principle 2) It Should Be Able to Assess the Impact of Data Protections on the Quality of Statistics

Under the Triple Trade-off, applying any SDL technique to mitigate disclosure risk will inevitably degrade the value of the resulting statistical product along one or both of the other dimensions (availability or quality). Effective management of this trade-off requires understanding and measuring what those impacts will be. Otherwise, how will an agency decision-maker assess whether the resulting statistical product's societal value is worth the release's marginal disclosure risk? Impacts on the availability dimension of the trade-off are typically quite straightforward to measure—the quantity and granularity of the statistics being released can be easily observed. However, the impact of SDL on the quality or accuracy of the resulting statistics can be much more difficult to effectively assess.

Part of the difficulty of assessing the quality of statistics protected by a DAS stems from the fact that no single or uniform measure of "quality" exists. Abstractly, one might posit that the relative *closeness* of the protected statistics to the unprotected data would be a good approximation of quality. But in the context of the uncertainty introduced through SDL, should that closeness be captured by measures of central tendency, the prevalence or magnitude of outliers, or some other assessment of whether a consumer of the resulting statistics would reach a different conclusion in their analysis?

Further complicating this assessment of post-SDL statistical quality is the fact that the degradation of quality caused by many SDL techniques (and the disclosure avoidance systems that employ them) often depends on the underlying characteristics of the data to be protected. Examples of this data dependency can be seen across a wide variety of SDL techniques. Suppression-based disclosure avoidance will have greater degradation on statistical products that have more frequent cell counts below the suppression threshold. The impact of top- and bottom-coding methods (a common form of coarsening) will largely depend on the prevalence of values above or below the recoding thresholds. Noise injection techniques that incorporate post-processing for error reduction or enforced consistency can also have complex or surprising effects on quality depending on the underlying sparsity of the data being protected.[1]

To be fair, disclosure avoidance systems that have data-dependent impacts on statistical quality can still permit comprehensive evaluation of the quality of the resulting statistics—that assessment would merely need to be performed *after* the protections have been applied. Reliance on *ex post* analysis of statistical quality, however, is inefficient within the framework of the Triple Trade-off. For any defined set of output statistics,

[1]The Census Bureau's recent iterative efforts tuning the 2020 Census Disclosure Avoidance System's TopDown Algorithm are an excellent example of the potential impact of this data dependency on statistical quality.(United States Census Bureau, 2023a)

agencies need to balance the degree of protection against the societal value of the resulting statistics. If the resulting value of the output is itself data dependent and cannot be effectively predicted in advance of the application of SDL, then the selected disclosure avoidance system would be an inefficient mechanism for implementing the chosen balance. The agency would either have to rely on an estimate of the impact (which represents an inherent imprecision in implementing the selected balance) or need to assess statistical quality on multiple applications of the system at various levels and configurations of protection (which would be resource inefficient).

In an ideal world, agency decision-makers should be able to assess the full impact of disclosure avoidance on statistical quality as part of their balancing of these dimensions under the Triple Trade-off. The more precisely a DAS can anticipate the impacts on quality before the application of SDL, the more effective and efficient that system will be at navigating the trade-off.

5.3 (Principle 3) It Should Be Able to Target Protection for Vulnerable Data Subjects

One of the unfortunate realities of disclosure limitation, particularly in the demographic and economic contexts, is that disclosure risk is not evenly distributed across the underlying population. The more a particular data subject's attributes or characteristics differ from those around them, the easier it often is to single them out or re-identify them in a statistical product. These "population uniques" have often been the primary focus of many disclosure avoidance systems, not just because they are typically the easiest to re-identify within statistical data products but also because the very characteristics that so easily differentiate them from their statistical neighbors can also expose them to greater potential harm should disclosure of their survey or census data occur.

Because of the increased vulnerability for these groups, an ideal DAS should be able to target increased protection, as needed, for these data subjects. A cynical interpretation of this principle might ask why some groups should deserve greater protection than others, but the reality is that disclosure risk is not evenly distributed across society. A uniform approach to disclosure avoidance that achieves any arbitrary (but uniform) level of disclosure protection will inherently over-protect some records, with the corresponding degradation of data quality or availability under the Triple Trade-off. Thus, equitable confidentiality protection can produce structured inequalities in the results (Bowen and Snoke, 2023). Finding an optimal balance along those three dimensions of the Triple Trade-off means being strategic and ethically minded about when and how more surgical protections should be applied and about how strong (or weak) those targeted protections need to be.

5.4 (Principle 4) It Should Be Able to Achieve Quality Targets for Statistical Products

Balancing statistical quality and availability against the disclosure risk of a statistical product under the Triple Trade-off relies mainly on the presumed societal value of the statistical product to be released. That value accrues from the use of those statistics. In situations where a single statistical product has many uses, it is often the case that some of those uses represent greater importance to society than others. Similarly, the relative impact of data protections on availability or statistical quality may vary depending on the particular use case. That is to say, some uses of a statistical product may be more (or less) sensitive to the impacts of SDL-induced uncertainty on accuracy or coverage. Efficient balancing of these dimensions under the Triple Trade-off means prioritizing the availability and statistical quality for those uses that represent the highest societal value and are the most sensitive to the impacts of SDL-induced uncertainty. An ideal DAS should give decision-makers the most flexibility to achieve and implement the agreed upon balance that reflects those priorities. To that end, much like an ideal disclosure avoidance system would allow decision-makers to target data protections for those most at risk or most likely to be harmed by disclosure (rather than applying uniform protections that will likely over-protect some records while under-protecting others), so too will the ideal DAS allow decision-makers to set and achieve statistical quality targets for those data uses with the highest societal value.

5.5 (Principle 5) It Should Track Cumulative Disclosure Risk Over Time

As noted above, every statistic that an agency releases that is derived from a confidential source carries a non-zero disclosure risk. While the risk from any single released statistic, especially at higher levels of aggregation, is often negligible, the cumulative risk across large volumes of data releases can be substantial. To effectively assess overall disclosure risk, it is important to assess the cumulative disclosure risk over these independent statistics as well as the potential interactions between them. Unfortunately, it is often these interactions between published statistics that complicate the cumulative assessment of disclosure risk. For example, two data tables protected using complementary cell suppression may each represent very low risks of disclosure on their own, but if a cell in one table can be used to infer the value of a suppressed cell in the other table, then the full set of primary and complementary suppressions can unravel through simple addition and subtraction. In this manner, the cumulative risk across both tables can be far greater than the sum of the individual disclosure risk of each table on its own. A DAS should be designed such that it is possible to assess the cumulative impact of each additional statistic, table, or product while accounting for the potential interactions of statistics (and protections) across each successive release.

5.6 (Principle 6) It Should Be Transparent

Statistical agencies have a professional and ethical obligation to be transparent about the known limitations of the statistics they produce (Committee on Professional Ethics of the American Statistical Association, 2022; United States Census Bureau, 2023b). Because any application of disclosure avoidance will have unavoidable consequences for the resulting usability of data, an ideal, applied DAS should provide meaningful transparency to those who will be using the resulting statistics. To that end, the DAS should be explainable to data users, including novice and intermediate users without formal training in disclosure avoidance. It should allow the uncertainty that is being introduced to protect confidentiality to be effectively assessed and for information or tools to be provided to enable data users to factor this uncertainty into statistical analyses. The system should also permit external assessment of the degree and scope of confidentiality protection and the correctness and efficiency of the system's implementation. In short, an ideal DAS will allow the agency to explain to policymakers and data users the impact that the system has on the availability and utility of the statistics.

5.7 (Principle 7) It Should Be Feasible

Statistical agencies operate within an environment of time-delimited research and production schedules and often of significant budgetary and human resource constraints. An ideal DAS will need to be able to effectively and efficiently operate within these constraints. Therefore, an ideal, applied system should support deployment, customization, integration, testing, and operation within reasonable time frames and with reasonable expenditure of resources.

6 Governance, Decision-making, and Parameter Selection

These principles are intended to be a tool for agencies to evaluate potential disclosure avoidance system options. However, they are not intended to prescribe the use of one particular approach or system over another. There is often tension between these principles. For example, the sophistication of an approach that can improve risk assessment and meet specific quality targets can also make transparency or feasibility more challenging.

Similarly, not all of these principles are necessarily equally important for a given statistical product. For a variety of reasons, an agency may place a higher value on transparency on the one hand or on protecting vulnerable groups on the other because they may consider one feature to be more critical for the organization at that particular moment.

Above all, these principles are intended to help differentiate decision-making about selecting a system (the particular SDL technique and algorithmic framework) from the implementation choices about the desired balance of the Triple Trade-off.

It should also be noted that throughout this discussion there have been repeated references to "agency decisions" and "agency decision-makers" without clarifying how these decisions are being made or who the agency officials are making those decisions.

Because of the technical complexities involved in disclosure risk assessment and mitigation, technical expertise in statistical disclosure limitation is an essential component of this decision-making. That said, risk assessment and mitigation is just one dimension of the broader Triple Trade-off, and considerations regarding the availability and quality of statistical products are also critically important to this decision-making. Agency governance of decision-making regarding the selection of a disclosure avoidance system and decision-making regarding the desired balance under the Triple Trade-off must, therefore, also reflect expertise in how the statistical products are intended to be used and the expected value to be gained from those uses. Finally, while the responsibility and authority for finding and implementing an appropriate balance under the Triple Trade-off ultimately rests solely with the agency itself, those who have to make these decisions will typically benefit from hearing the perspectives and viewpoints of the agency's external stakeholders.

7 Conclusion

Selecting a DAS is a challenging task. The choice of SDL techniques and the algorithmic framework for their deployment will typically have substantial consequences for the resulting system's ability to meet the statistical agency's objectives and requirements regarding the availability, accuracy, and confidentiality protection of its statistical products. The seven principles outlined above, which help identify the characteristics that such a system would ideally have, do not purport to state that any particular disclosure avoidance approach or system is inherently better or worse than another. Instead, they are intended to provide a framework for a statistical agency to use when evaluating between candidate systems that will allow agencies to determine, based on the agency's current priorities and objectives, which system will best assist the agency in producing and releasing high-quality statistical products while also protecting the confidentiality of their data subjects.

References

Abowd, J. M. and Hawes, M. B. (2023a). 21st century statistical disclosure limitation: Motivations and challenges. https://www.census.gov/library/working-papers/2023/adrm/ced-wp-2023-002.html. Working Paper ced-wp-2023-002.

Abowd, J. M. and Hawes, M. B. (2023b). Confidentiality protection in the 2020 us census of population and housing. *Annual Review of Statistics and Its Application*, 10:119–144.

Bowen, C. and Snoke, J. (2023). Do no harm guide: Applying equity awareness in data privacy methods. Technical report, Urban Institute.

Committee on Professional Ethics of the American Statistical Association (2022). Ethical Guidelines for Statistical Practice. https://doi.org/10.5281/zenodo.7092386. Accessed: 2022-09-13.

Dinur, I. and Nissim, K. (2003). Revealing information while preserving privacy. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '03, page 202–210, New York, NY, USA. Association for Computing Machinery.

Hawes, Michael B. (2021). Promoting Data Access and Data Protection: How emerging technologies can help agencies navigate the triple constraint of data. https://www.bea.gov/system/files/2021-03/Michael-Hawes.pdf. Presentation to the Federal Advisory Committee on Data for Evidence Building, March 19, 2021.

United States Census Bureau (2023a). U.S. Census Bureau Demonstration Data Product Suite (2023-04-03). https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/newsletters/new-2010-redistricting-dhc-demo-data.html.

United States Census Bureau (2023b). U.S. Census Bureau Statistical Quality Standards. https://www2.census.gov/about/policies/quality/quality-standards.pdf.