Disclosure Avoidance and the Census

Select Topics in International Censuses[1]

By Sam Dupre

Released October 2020

INTRODUCTION

One of the most important roles that national statistical offices (NSOs) play is to carry out a national population and housing census. In so doing, NSOs have two data stewardship mandates that can be in direct opposition. Good data stewardship involves both safeguarding the privacy of the respondents who have entrusted their information to the NSOs as well as disseminating accurate and useful census data to the public. If an NSO wants the public to participate and provide accurate responses during a census, then the public must believe that their responses will be held securely.

This technical note discusses the three types of disclosure events, forms of attacks that may occur, and the four—generally sequential—disclosure avoidance phases that an NSO could go through in a census to ensure good data stewardship.

KEY CONCEPTS

Data Types

There are two types of data that an NSO might release to the public: microdata and aggregate data. Microdata are the set of responses for a unit of observation (a household in the case of a population and housing census). Microdata are usually released to the public as a small sample of all respondent data. Aggregated data are summary information for entire groups of individuals in the form of frequency counts or magnitude figures (e.g., means, ranges, or other summary statistics).

To maintain public trust, an NSO should not release the full, unedited census dataset until enough time has passed that privacy is no longer pertinent (UNECE-CES, 2015). As an example, the United States only releases census data files after 72 years (USCB, n.d.).

Forms of Disclosure

There are three primary forms of disclosure. Each one represents a different level of risk to respondents (Buron and Fontaine, 2018; UNSD, 2015). *Identity disclosure* occurs when respondent identity is directly linked to a disseminated data record (e.g., through name, address, identification number, fingerprint, e-mail address, or a telephone number). *Attribute disclosure* occurs when values in disseminated data disclose other attributes of an individual. *Inferential disclosure* (along with subsequent reconstruction and reidentification) occurs when disseminated data are used to infer values for specific respondents based on statistical properties of the released data. The occurrence and likelihood of disclosure differs depending on the dissemination product (McKenna and Haubach, 2019). As an example, a report stating that all members of an ethnic group live in a single geographic region has a high risk of attribute disclosure but a low risk of identity disclosure. This is because an attacker could determine the region in which a person lives if that attacker knows that a person is a member of that ethnic

[1] This technical note is part of a series on *Select Topics in International Censuses* (STIC), exploring matters of interest to the international statistical community. The U.S. Census Bureau helps countries improve their national statistical systems by engaging in capacity building to enhance statistical competencies in sustainable ways. Any views expressed are those of the author(s) and not necessarily those of the U.S. Census Bureau.

U.S. Department of Commerce
U.S. CENSUS BUREAU
census.gov

United States Agency for International Development
www.usaid.gov

United Nations Population Fund
www.unfpa.org

Box 1.

Forms of Attack

Steps in an Attack

Reidentification: A record is matched to a human source.

Reconstruction: Anonymized values for each record are then deanonymized.

These steps are not necessarily sequential, as reconstructed values could be used to reidentify sources or vice versa.

Common Attacks

Database reconstruction attack: Matching a field (the "key") in an anonymized dataset with a common field in a public dataset.

Tracing attack: Attempting to find a target's data within a dataset. Depending on the comprehensiveness of the dataset or the sensitivity of the topic, even just identifying if a person is included could represent a serious breach of privacy or create the opening for further reconstruction.

Differencing attack: One of the biggest risks with spatial data. Works through using differences in repeated queries to learn information about records by comparing subsets.

Note: Information compiled from Dwork et al., 2017; McKenna and Haubach, 2019; and UNSD, 2015.

group. The methods of attacks vary, but some of the most common are included in Box 1.

THE DISCLOSURE AVOIDANCE PROCESS

The disclosure avoidance process generally has four phases: (1) Risk Assessment, (2) Public Consultation, (3) Disclosure Controls, and (4) Archival and Access/Release. The phases begin prior to the census and continue until after the main census timeline is complete. Table 1 details these four phases and their subphases.

NSOs employ three strategies to safeguard respondent privacy throughout each of these phases. They *restrict collection*, minimizing the collection of data on sensitive topics where possible (NASEM, 2017; UNSD, 2015). They *restrict data*, controlling which components of the data are released and in what form (e.g., aggregate data or microdata), and what statistical controls are applied. Finally, they *restrict access*, controlling which users can access data along with their degree of access.

Table 1 presents an overview of these phases. This is not meant to detail all technical minutiae, but instead to provide a primer on technical considerations for the process. Due to the complexity and profound importance of these activities, following this overview we will include additional information on post-enumeration risk assessment, statistical control phases, and archival access or release.

Table 1.

Disclosure Avoidance Phases

Phase	Subphase	What this entails
Risk Assessment	Internal Assessment	Early in census planning, review risk based on the type and sensitivity of data to be collected and released.
	External Assessment	Following internal assessment—but prior to enumeration—hire external consultants to conduct an independent risk assessment.
	Second Internal Assessment	Following enumeration and disclosure controls, repeat internal risk assessment including quantitative review of collected data.
	Disclosure Review Board	Before, during, and after the census, have a review board conduct risk assessments for new data dissemination plans. This board should rereview disseminated products as technology improves.
Public Consultation	N	Before the census, consult with stakeholders on their privacy concerns and data needs. Data needs cover what data they want released and in what format. Use this information to guide NSO risk assessment from the start of the planning process and target historically hard-to-count populations (see the U.S. Census Bureau's guide to Counting the Hard to Count in a Census [2019a]). Box 2 presents a case study on how the United Kingdom Office of National Statistics addressed the public consultation phase.
Disclosure Controls	Legal Controls	Before the census, establish legal statutes that mandate NSO responsibility to safeguard respondent data, specifically laying out how data may be released. Doing so provides legal backing that sanctions NSO decision-making.
	Physical Controls	Before enumeration, establish policies for material disposal, facility access, and how a retained representative sample will be treated. Disposal includes paper forms and Computer-Aided Personal Interviewing device data wipe.
	Technical Controls	Before enumeration, set policies which prevent online census response interception, secure lost Computer-Aided Personal Interviewing device data, enforce NSO network security, and control staff access to respondent data.
	Statistical Controls	After enumeration, apply statistical measures to respondent microdata (pre-tablular) or to aggregate data (post-tabular). Specific measures used depend on what form of release is planned.
Archival Access/ Release	N	After the census, archive microdata (raw files and edited files after statistical controls), metadata, and paradata and make them available to stakeholders. Refer to the U.S. Census Bureau's Census Data Archiving and Preservation guide for detailed for detailed information on secure data archival (2019b).

N Not applicable.
Note: Information compiled from Lauger et al., 2014; McKenna and Haubach, 2019; NASEM, 2017; UNECE-CES, 2015; and UNSD, 2015.

Additional Information on Post-Enumeration Risk Assessment and Statistical Controls

Risk assessment and statistical controls tend to be some of the most technically complex aspects of disclosure control (Table 1). The specific measures to use depend on:

- The form of release (e.g., Public-Use Microdata Samples [PUMS] versus aggregate data) (McKenna and Haubach, 2019).
- The level of detail planned (a PUMS file with coarse data aggregation might require a minimum population of 100,000, while one with highly detailed data might require a minimum population of 400,000) (Buron and Fontaine, 2018).

Box 2.

Case Study: The Office for National Statistics (ONS) of the United Kingdom

In preparation for the 2021 Census, between 2015 and 2018, ONS sought multiple rounds of public consultation on topics and demands from the 2021 Census. Following each round, the ONS published detailed information on (1) initial plans, (2) public responses, (3) ONS plans based on those responses, and (4) on the implications of changing plans for equal representation in the 2021 Census.

Source: ONS, 2018.

Following enumeration, NSOs could repeat risk assessment based on the data collected—factoring in updated release plans—and apply statistical control measures. This integrated five-step process is as follows:

Step 1: Eliminate Personally Identifying Information (PII)

Remove direct identifiers such as name, address, and any government identification numbers from records to prevent direct identity disclosure (UNSD, 2015).

Step 2: Identify Sensitive Records, Cells, and Categories

While statistical disclosure risk assessment requires qualitative subject matter expertise to identify locally sensitive/vulnerable topics and groups (UNSD, 2015), there are quantitative measures to assess disclosure risk. Using quantitative measures allows for clear comparison between different dissemination options and provides defensible legal backing to NSO decision-making (NASEM, 2017).

Table 2 presents a series of common challenges that an NSO could encounter in identifying sensitive records, cells, and categories along with guidance on methods for quantitatively assessing if that sensitive item should be flagged for statistical control measures.

Step 3: Address the Risk

Statistical controls may be *perturbative* or *nonperturbative* (Antal et al., 2017). Perturbative measures slightly alter the data in controlled ways, changing data structure as minimally as possible. Nonperturbative measures work by removing (or aggregating) table cells, geographic areas, or data records that meet certain levels of risk. Perturbative methods tend to preserve data structure more reliably and suffer less information loss than Nonperturbative methods (Antal et al., 2017).

Table 2.
Common Features Leading to Sensitive Records, Cells, and Categories

Challenge	Why this is a challenge	Assessing this risk quantitatively
Cells with small counts exist.	The risk of identity disclosure increases when there are very few records within a grouping.	Flag all units that fall below a standard threshold. For the U.S. Census Bureau's American Community Survey (ACS) and the 2010 Census PUMS file: • Each category of a categorical variable must contain at least 10,000 unweighted people or households. • All geographic areas (including urban/rural status) must contain at least 50 unweighted people or households for a single variable. • Tabulations require a mean cell size of at least three unweighted cases.
Nonzero counts exist for sensitive groups.	Even knowing that people with certain characteristics exist could lead to privacy disclosure.	Flag all cells for preidentified sensitive characteristics or combinations of characteristics.
Different subsets of results include the same population(s).	These subsets could be compared in a differencing attack to infer respondent data.	Flag nonnested geographies and respondent groupings for further review, paying special attention to any cases where only small differences exist between the repeated subsets for a population.
Individuals within a household are flagged as at risk.	If an individual presents a disclosure risk, then the entire household could be at risk.	Assess risk at the individual level for each variable and geographic level. Aggregate individuals to form households and flag any household with an at-risk member.
Outliers exist within the responses for any variable.	Those who are at the top or bottom of the response distribution are easier to identify compared to those with responses closer to the mean.	For continuous variables, flag records with values near the maximum and minimum of the distribution. Typically, this would include the bottom/top 0.5 percent of all values (or 3 percent of all nonzero values if that would include more records).
	The respondent(s) with the largest reported values in a grouping are most likely to be at risk from the other respondents in the extremes of that grouping.	The (n, k) rule, the p% rule, and the pq rule flag cases where outlier respondents might be able to identify other outlier respondents based on their own responses. The (n, k) rule flags a variable if the values for the largest n respondents makes up at least k percent of the total values. The p% rule and the pq rule flag cases where other respondents could estimate the values for the respondent with the largest value to within p percent of the true value. The specific values that an NSO uses for n, k, p, or q are generally confidential as even this information could make data vulnerable to an inferential disclosure attack.

Note: Information compiled from Antal et al., 2017; Buron and Fontaine, 2018; Lauger et al., 2014; McKenna and Haubach, 2019; and OECD, 2005.

Nonperturbative

Primary and secondary/complementary suppression. Primary suppression protects against identity/attribute disclosure by replacing cells or records with a marker that identifies they have been suppressed or show as "No Data" (Antal et al., 2017). Secondary suppression involves suppressing additional nonflagged cells so that suppressed values cannot be derived through inferential disclosure. Alternatively, all problematic variables or entire flagged groups or geographies could be suppressed from dissemination (UNECE-CES, 2015).

Recoding. When there are too few records for a value or range of values, it may be combined with other groups, records, columns, or rows until any threshold is met. When publicly available data could be linked to census data, recoding may be necessary to prevent attribute or inferential disclosure even when a standard threshold is already met. Options for recoding quantitative data include rounding, interpolation within a predefined range/distribution, or reduction to quantiles to reduce data specificity (Dajani et al., 2017). Top-coding and bottom-coding are forms of recoding used to disguise outliers for continuous variables. Outliers in the top or bottom of a percentile threshold are replaced with the cutoff value or with the mean or median of all top/bottom-coded values.

Perturbative

Noise addition. Random noise is added to risky cells by making small changes to original values. The noise added ideally maintains data structure for that variable by controlling for bias, variance, numerical frequencies, and adjusting zero-value cells (Antal et al., 2017).

Record swapping, rank swapping, and shuffling. Record swapping involves matching pairs of records on some criteria and then swapping nonequal values between those pairs (Antal et al., 2017). The parent geography of the swapped pairs should be the same whenever possible, in order to minimize data disruption and minimize geographic shift (e.g., swap within regions, but not between regions) (Buron and Fontaine, 2018), though this is not always the case when substantial disclosure risk exists (Lauger et al., 2014). Rank swapping and shuffling switch values for some variable among records that have similar values for that variable.

Synthetic data. Create a statistical model that describes the dataset and then replace unique records with a modeled value (Dajani et al., 2017). These datasets also require risk assessment as disclosure incidents could still occur; however, they let researchers access data that otherwise might present a prohibitive risk.

Step 4: Check Results

Check retained risk using the measures described in Step 2 and check the degree of information lost (e.g., increased variance in parameter estimation or introduced bias). To assess information loss, check:

- If any perturbation substantially changed minima/maxima, mean/median/mode, or percentiles (absolute and relative differences) (Antal et al., 2017).
- The proportion of cells where perturbation exceeds a prespecified degree of change, as small changes in low-density areas could have larger effects than relatively much larger changes in high-density areas (Buron and Fontaine, 2018).
- If perturbation adds a substantial percentage of "false positives" (zeroes perturbed to nonzeroes) and "false negatives" (nonzeroes perturbed to zeroes) (Buron and Fontaine, 2018).
- If data relationships still exist after perturbation (e.g., expected parity or inequality—or a specific statistical relationship—exists between two variables) (Antal et al., 2017).

Step 5: Conduct Internal Attack Studies

NSOs can conduct internal disclosure attack studies to uncover new vulnerabilities as novel threats emerge (McKenna and Haubach, 2019). These attack studies should use the same methods and technologies an attacker might, incorporating public and private datasets and new technological developments. These tests could be applied to both new and old data releases to ensure that previously anonymized records have not become vulnerable to disclosure.

Additional Information on Archival and Access/Release

Microdata (both raw files and edited files following statistical disclosure avoidance), metadata, and paradata could all present risks for data disclosure singly or with other sources (UNECE-CES, 2015). The NSO should retain original unedited versions of the data internally and create a log of all edits, stored separately from the anonymized data files (Van den Eynden et al., 2011). On occasion, a representative sample of completed census forms may be retained by the NSO. If so, all data privacy principles could be applied before any release may occur.

Some arrangements that NSOs could use to maintain secure forms of access include data enclaves or remote access facilities, online databases to request datasets or analyze data, licensing arrangements for verified

users, or the public release of a PUMS file (FCoSM, 2005; Hundepool et al., 2012). Irrespective of the arrangement in place, common principles apply. For security reasons, NSOs should not allow external users to access internal networks. Data not approved for public release could be encrypted prior to any transfer off internal, secured networks. Data enclaves should have no access to the Internet, external networks, or USB ports (UNSD, 2015). Additional disclosure avoidance measures should automatically trigger when similar tables are requested multiple times, as a safeguard against differencing attacks through repeated subsetting. Arrangements should involve unannounced audits of data storage facilities, review of statistical outputs, and cover the disposal of data and derived files. In order to be effective, all arrangements could be legally binding and contain penalties for violation (FCoSM, 2005; UNSD, 2015).

CONCLUSION

The value of a census is severely diminished without dissemination of timely, usable data. However, increasing detail in data releases escalates the risk that respondent privacy could be violated. This risk is growing in the age of Big Data, as developments in data mining tools, data georeferencing, and statistical data processing capabilities increase the likelihood of data disclosure incidents. NSOs can meet their public mandates by managing risk using the policies and procedures introduced in this note.

REFERENCES

Antal, L., T. Enderle, and S. Giessing, *Harmonised Protection of Census Data in the ESS: Statistical disclosure control methods for harmonised protection of census data*, Eurostat Centre of Excellence on Statistical Disclosure Control, The Hague, 2017.

Buron, M. L., and M. Fontaine, Confidentiality of Spatial Data, in Loonis, V. and Marie-Pierre de Bellefon, *Handbook of Spatial Analysis: Theory and application with R*, chapter 14, Insee Méthodes No. 131, Eurostat, The Hague, 2018.

Council of Europe (CoE), *Guidelines on the Protection of Individuals with Regard to the Processing of Personal Data in a World of Big Data*, Directorate General of Human Rights and Rule of Law, Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Strasbourg, France, 2017.

Dajani, A.N., A.D. Lauger, P.E. Singer, D. Kifer, J.P. Reiter, A. Machanavajjhala, S.L. Garfinkel, S.A. Dahl, M. Graham, V. Karwa, H. Kim, P. Leclerc, I.M. Schmutte, W.N. Sexton, L. Vilhuber, and J.M. Abowd, The Modernization of Statistical Disclosure Limitation at the U.S. Census Bureau, in *September 2017 Meeting of the Census Scientific Advisory Committee*, Suitland, MD, 2017.

Dwork, C., A. Smith, T. Steinke, and J. Ullman, *Exposed! A Survey of Attacks on Private Data*, Annual Review of Statistics and Its Applications, 4(12): 1–24, 2017.

Federal Committee on Statistical Methodology (FCoSM), *Statistical Policy Working Paper 22: Report on Statistical Disclosure Limitation Methodology Version 2*, U.S. Office of Management and Budget, Washington, DC, 2005.

Hundepool, A., J. Domingo-Ferrer, L. Franconi, S. Giessing, E. Shulte Nordholt, K. Spicer, and P.P. de Wolf, Statistical Disclosure Control, In: *Wiley Series in Survey Methodology, Wiley*, Chichester, United Kingdom, 2012.

Lauger, A., B. Wisniewski, and L. McKenna, Disclosure Avoidance Techniques at the *U.S. Census Bureau: Current practices and research, research report series (Disclosure Avoidance #2014-02)*, Center for Disclosure Avoidance Research, U.S. Census Bureau, Washington, DC, 2014.

McKenna, L. and M. Haubach, *Legacy Techniques and Current Research in Disclosure Avoidance at the U.S. Census Bureau*, Research and Methodology Directorate, U.S. Census Bureau, Washington, DC, 2019.

National Academies of Sciences, Engineering, and Medicine (NASEM), *Innovations in Federal Statistics: Combining data sources while protecting privacy*, The National Academies Press, Washington, DC, 2017, <https://doi.org/10.17226/24652>.

OECD, Glossary of Statistical Terms, <https://stats.oecd.org/glossary/detail.asp?ID=6943>, 2005, accessed on July 15, 2020.

Office for National Statistics (ONS), Initial View on 2021 Census Output Content Design: Response to consultation, Office for National Statistics, United Kingdom, 2018.

United Nations Economic Commission for Europe—Conference of European Statisticians (UNECE-CES), *Recommendations for the 2020 Censuses of Population and Housing*, United Nations Publications, New York, NY, 2015.

United Nations Statistics Division (UNSD), *Principles and Recommendations for Population and Housing Censuses, Revision 3*, United Nations Publications, New York, NY, 2015.

United States Census Bureau, *Counting the Hard to Count in a Census*, Select Topics in International Censuses, <www.census.gov/content/dam/Census/library/working-papers/2019/demo/Hard-to-Count-Populations-Brief.pdf>, 2019a.

_____ , *Census Data Archiving and Preservation*, Select Topics in International Censuses, <www.census.gov/content/dam/Census/library/working-papers/2019/demo/Archiving-Brief.pdf>, 2019b.

_____, The "72-Year Rule," <www.census.gov/history/www/genealogy/decennial_census_records/the_72_year_rule_1.html>, n.d., accessed on July 15, 2020.

Van den Eynden, V., L. Corti, M. Woollard, L. Bishop, and L. Horton, *Managing and Sharing Data*, UK Data Archive, UK, 2011.

The Select Topics in International Censuses (STIC) series is published by International Programs in the U.S. Census Bureau's Population Division. The United States Agency for International Development sponsors production of the STIC series, as well as the bilateral support to statistical organizations that inform authors' expertise. The United Nations Population Fund collaborates on content and dissemination, ensuring that the STIC series reaches a wider audience.