Glossary

This glossary provides explanations of how particular technical terms are used in the context of statistical confidentiality. Terms generally used in broad fields of statistics, information technology or empirical research are not included. Nor do we include broad terms such as privacy and confidentiality, which are themselves major subjects of the whole of this book.

Additive table A contingency table in which the interior cells sum to the associated marginal totals. Tables can become non-additive when some perturbative disclosure limitation techniques are applied to them.

Aggregate data Data which summarize information across a population (or sample). Summary statistics, frequency tables, and magnitude tables are all examples of aggregate data.

Analysis server An Internet or LAN-based remote access system enabling the analysis of data sets to be conducted remotely.

Analytical completeness The extent to which a disclosure-limited data release allows the same analyses to be conducted as prior to the disclosure limitation being applied.

Analytical validity The extent to which a disclosure-limited data release leads to the same statistical inferences as prior to the disclosure limitation being applied.

Anonymization A process to make practically impossible the identification of subjects in a database, that is, a process of making the risk of identity disclosure negligible.

A priori knowledge Knowledge that a data snooper has which would facilitate attempts to disclose information about individual population units. This might be in the form of pre-existing knowledge about the population units, knowledge of the disclosure control process that has been applied to the data set, or response knowledge.

G.T. Duncan et al., *Statistical Confidentiality*, Statistics for Social and Behavioral Sciences, DOI 10.1007/978-1-4419-7802-8,

Synonym: External knowledge

Attribute suppression Statistical disclosure limitation by eliminating all or some values of one or more attributes.

Attribution The association of information in a data set with a particular population unit.

Attribute disclosure The disclosure of information about a **population unit** without (necessarily) the identification of that population unit within a data set. This typically refers to aggregate data where something can be inferred about individuals who possess a certain combination of characteristics.

(Disclosure) Auditing The process of checking whether processed data are protected or not. When the data are tabular, then this process is done by computing the minimum and maximum value of all the risky cells. On large tables the computation of these two extreme values may be through integer linear programming.

Cell perturbation A statistical disclosure limitation method for tabular data whereby the values of some (or all) cells are replaced by some value in an interval.

Cell suppression A disclosure limitation method for tabular data whereby the values of particular cells are not released.

Confidentiality promise An assurance given to respondents that their data, once provided, will not be disseminated in a form in which they are identifiable or through which information can be learnt about them.

Confidentiality breach A situation where a confidentiality promise is broken, especially when a data snooper has actually identified a subject in the database and made use of this information.

Consent The process through which a respondent gives permission for data about them, collected by a data stewardship organization, to be used in a given way.

Controlled rounding A disclosure limitation method for tabular data whereby the values of all cells are replaced by a value in a finite set. Typically, for each cell, this set contains only two values, which are the numbers rounding up and down the original cell value to a multiple of a given base number (for example, 10). Marginal totals are maintained.

Controlled tabular adjustment (CTA) A disclosure limitation method for tabular data whereby the values of all cells are replaced by different values. It is a particular method of cell perturbation. CTA adjusts cell values to satisfy disclosure limitation ranges and tabular constraints, such as additivity, while minimizing data loss as measured by some linear measure of overall data distortion, such as the sum of the absolute values of the individual cell value adjustments. CTA replaces each risky cell by either of the two endpoints of its protection range. Certain non-risky cell values are adjusted by small amounts to restore additivity.

Cyclic perturbation A disclosure limitation method for tabular data whereby the values of all cells are replaced by different values. It is a particular technique of cell perturbation in which patterned collections of four or more cells, called data cycles, are randomly modified with some cell values increasing by one and others decreasing by one, in such a way that each row and column sum is undisturbed.

Data dissemination Any process through which (access to) a set of data are given to data users.

Data divergence The difference either between two data sets or a data set and the world in how information about a given population unit is recorded. This can be caused by coding differences, response errors, data aging, data entry errors, and other forms of misclassification.

Data protection A broad term which covers a set of legal and ethical principles and instruments which collectively delineate the ways in which data may be fairly processed by data stewardship organizations. Data protection operates at national and international levels.

Data snooper An individual, group, or organization that seeks to identify individual population units within a data set and/or discover information about a population unit, usually through a statistical linkage process of information already known to information contained within the data set.

Data Stewardship Organisation (DSO) An organization which captures, processes, holds, and/or disseminates data about population units. A DSO is under legal or ethical obligations to maintain confidentiality and provide data releases of high utility to users.

Data swapping An SDL method in which certain attribute values for some records in the source data are exchanged. For example, a sample of households may be selected and matched on a set of selected key variables with households in nearby geographic areas that have similar characteristics (such as the same number of adults and same number of children). Values of an attribute such as household income may be swapped.

Data user Any person using a data set for a bona fide, typically statistical, purpose.

Data utility The usability or research value of a given set of data. In the context of disclosure-limited data, high data utility requires both analytical completeness and analytical validity.

Data-world divergence The difference between a data set and reality. Sources of data divergence include data outdating, response errors, coding or data entry errors, differences in coding, and the effects of disclosure limitation.

Deidentification The removal of direct identifiers from a data set or individual record.

Direct identifiers Information that can identify a data subject directly, such as name or social security number. Also includes combinations of attribute values, such as address and age, that uniquely identify a data subject.

Disclosure auditing See auditing.

External knowledge Knowledge that a data snooper may have that is external to that in a particular data release.

Global recoding Any method of recoding the categories in a whole data set to a smaller set of categories.

Key variables A combination of attributes that increase disclosure risk because a data snooper may link values to those in an identified external data set.

Hierarchical data Used particularly to describe microdata which has grouping variables for two or more levels of analysis. The vast majority of microdata is geographically hierarchical. Another variant is hierarchical household data which contain all members of a set of sampled households.

Identification The association of a population unit whose identity is known with a particular microdata record.

Identification file A database, data set, or other form of data storage which contains formal identifiers for individual population units.

Identification risk The probability that a data snooper can identify a data subject in a released data product.

Identity disclosure Identity disclosure occurs when a data subject is identified from released data.

Integer linear programming Mathematical tools for optimization. Typically there is a linear objective function that must be maximized or minimized by choosing values of specified variables. These variables are subject to linear systems of equalities and/or inequalities with integer variables and possibly continuous variables.

Interval publication A disclosure limitation method for tabular data whereby the values of particular cells are replaced by intervals of values. For each value in each interval there should exist values in the other intervals such that together with the values which have not been replaced the resulting table is coherent. Each interval should contain the original value that it is replacing.

Linking The process of matching a data subject in released data to an identified subject in an external database.

Local suppression A disclosure limitation method where particular data units within a microdata file are coded as missing. This method is used within the **ARGUS** system.

Lower bound The minimum possible value of a given cell in a table of aggregate values which has been perturbed so that the exact value is not given (and possibly is even suppressed).

Macrodata Summary data, most especially data released in the form of a contingency table, hence with categorical attributes.

Magnitude tables Tabular data where each cell value has been obtained by adding the response value of all the contributors that fit within the categories of the key variables describing such cell. The response variable may be continuous or integer numbers.

Marginal total A cell value that has been computed by summing other cell values.

Matrix masking SDL methods that transform an $n \times p$ (cases by variables) data matrix Z through pre- and post-multiplication and the possible addition of noise.

Masking Any SDL method that prepares data for release by stochastic or deterministic transformation of the source data.

Mathematical programming An area of mathematics devoted to study mathematical models and algorithms to solve optimization problems. Integer linear programming is a sub-area, but there are other areas like convex optimization and quadratic programming.

Microaggregation An SDL method in which records are aggregated into groups. Instead of releasing the actual values for individual records, the mean (typically) of the group is released. Confidentiality is protected by each group having at least a minimum number of observations.

Microdata Data composed of records on individual data subjects. Each record might refer to an individual person, household, business, or other entity. The data may be directly collected for statistical purposes or obtained from other sources, such as administrative sources.

Minimal sample unique A combination of values within a data set that is unique within the data set and for which no subset of values is also unique.

Multivariate additive noise A statistical disclosure limitation method in which the released data are created by adding a random vector of disturbances to the source data. The multivariate distribution of the random vector is chosen to lower disclosure risk while maintaining adequate data utility.

Noise addition A statistical disclosure limitation method in which the released data are created by adding a random disturbance to some or all values in the source data.

***n* rule** A procedure to detect the risky cells (those requiring protection) in a table, so a procedure to solve the so-called primary problem. This procedure is a simple mechanism that takes into account only the number of contributors to each cell. Given a threshold value n, a cell is classified as "risky" when the number of contributors to this cell is less than or equal to n.

Parallel divergence A situation where two data sets both contain the same value for a given population unit/variable but both differ from a veridical representation of that population unit/variable. For example, somebody who

lies to the tax collector about their income may well also lie to the census agency.

Perturbation Any disclosure limitation process where values for given data units are changed either systematically or randomly.

Primary problem Checking whether there is something to protect or not in a dataset. When there is something that needs protection, it must be individuated. For example, given a table, the primary problem is the problem of finding the cells (if any exist) that will need protection. The cells individuated by solving the Primary Problem are called "Primary Cells" (or risky cells). The secondary problem is protecting the risky cells.

Poisson model A model based on the assumption that the counts in a contingency table are Poisson distributed.

Population uniqueness The proportion of population units, within a given population, which has a unique combination of values, for a given set of variables.

Population unit A socioeconomic entity about which data may be collected and which is a member of a set of such entities. The set is referred to as the population. Entities may be individuals, households, families, businesses, or other organizations.

***p*/*q* rule** It is a procedure to detect which cells require protection in a table. In other words, it is a procedure to solve the so-called primary problem. This procedure is a mechanism more elaborated than the so-called *N* rule. In particular, the *p*/*q* rule takes into account the number of contributors to each cell and also the value of the major contributors.

Post-tabular SDL Statistical disclosure limitation processes applied to dataset once it has been aggregated into tabular format.

Pre-tabular SDL Statistical disclosure limitation processes applied to dataset prior to it being aggregated into tabular format.

Primary suppressions The cells individuated by a rule to solve the Primary Problem on a table. They are called suppressions when the method to solve the Secondary Problem is the cell suppression technique.

Probabilistic record linkage Record linkage processes where non-exact linkages are allowed.

Random rounding A statistical disclosure limitation method for tabular data that rounds each cell of a table up or down to a multiple of a given base number (say 5). It makes use of a probability function to decide the direction of the rounding for each value. For example, a cell value i will go down with probability $(1-i/5)$ and up with probability $i/5$. The main advantages of random rounding are the simplicity of the approach and the unbiased feature of the resulting table. The main disadvantage is that the resulting table may not be additive when the marginal cells are also modified with the same approach.

Random unique A record which is not special unique on a given set of variables.

Record linkage The process of linking records on different data sets. This can be done by researchers for bona fide analytical reasons or by a data snooper attempting to link identification information to an anonymized data set.

Record suppression A disclosure control process whereby whole records are removed from a data set before dissemination (or, in the cases of samples from census data, deliberately not sampled).

Reidentification The process of determining the identity of data subjects in data releases that have been subject to statistical disclosure limitation, that is, have been deidentified.

Research Data Center A physical facilility maintained by a data stewardship organization where data users may make use of the data for statistical purposes under restricted access conditions.

Response knowledge Knowledge that a particular population unit is represented in a given data set. This usually is in the form of knowledge that the population unit responded to the survey. Response knowledge in respect of a sample survey data set can greatly increase the risk of disclosure.

Restricted access An approach to disclosure limitation in which data access is controlled by the data stewardship organization. Access controls can be physical, technical or administrative.

Restricted data An approach to disclosure limitation in which the data are transformed, reduced, or otherwise masked and could be used on its own or in combination with restricted access.

Risk metrics Numerical indices (possibly estimated probabilities of identification or attribution) designed to indicate the likelihood of disclosure events under a given set of assumptions.

Risky cells Cells in a table that have been deemed to have high disclosure risk.

Rounding A method of disclosure limitation whereby values are rounded to a particular base. This is usually applied to frequencies within a table of counts.

R-U confidentiality map A graph of the trajectory in (disclosure risk, data utility) space as the extent of statistical disclosure protection of a given method increases.

Safe setting A form of restricted access where the data analysis environment is highly controlled. This will typically be within the offices of the data stewardship organization or in satellite offices, such as research data centers, where it has assurance that confidentiality standards are maintained.

Sample uniqueness The proportion of records, within a sample data set, which have a unique combination of values, for a given set of variables.

Sensitive cells See risky cells

Secondary problem The problem that must be solved after the Primary Problem has been solved and has generated a set of risky cells. The Secondary Problem consists of protecting the risky cells. The way to proceed depends on the chosen method (like cell suppression and controlled rounding).

Secondary suppressions Used in disclosure limitation for aggregate data these are cell suppressions that are necessary in order to stop a data snooper (or data user) from recovering the original cell values. They are the output of the Secondary Problem when the chosen method is cell suppression. *"Complementary suppressions" is a synonym.*

Sensitivity A measure, using qualitative and often subjective, of the damage that would be caused (usually to the respondent concerned) if a given piece of information is disclosed about either a given population unit or an imaginary "typical" population unit.

Social acceptability A conceptual component of sensitivity which considers how responses to surveys might vary in their compliance with social norms and expectations. Consider, for example, a survey of sexual practices or attitudes toward political extremism.

Special unique A record within a microdata set which is sample unique on a given set of variables and also on a subset of those variables.

Special unique detection algorithm A method for calculating and grading the disclosure risk for individual-level categorical microdata by determining the number and scale of special unique combinations of variables for each record within a data set within some defined key.

Statistical disclosure The process by which information is inferred about individual population units.

Statistical disclosure limitation (SDL) The process of lowering the disclosure risk of data releases intended for statistical analysis purposes. *"Statistical disclosure control" is a synonym.*

Subtraction A method of attacking a table-aggregated dataset (typically a table of counts) by removing population units for which data are already known for the dimensions of the table and thereby reducing the table to a residual, possibly disclosive, form.

Swapping key A set of key variables used for pairing records in record swapping disclosure limitation regimes.

Swapping partner Records within a data swapping system which are paired on the basis of having the same (or possibly similar) values on a swapping key.

Synthetic data Data to be released which have been probabilistically generated based on a model estimated from the source data.

Target variables Variables within a data set deemed likely to be of interest to an intruder. Target variables will generally be (i) unavailable to an intruder and (ii) sensitive in some way.

Topcoding An SDL method used with interval-scale or many-category ordinal variables where values above a certain threshold are revealed only in aggregate form. Age and income often are topcoded, with only an indication that the values are above the threshold.

Unrestricted access A data dissemination process whereby no limits are placed on access nor is such access monitored. This now is mainly used for the delivery of aggregate statistics over the web.

Upper bound The maximum possible value of a given cell in a table of aggregate values which has been perturbed so that the exact value is not given (and possibly is suppressed).

References

Abowd, J.M., Woodcock, S.D.: Disclosure limitation in longitudinal linked data. In: Doyle, P. et al. (eds.) Confidentiality, Disclosure, and Data Access, pp. 135–166. North Holland, Amsterdam (2002)

Abowd, J.M., Woodcock, S.D.: Multiply-imputing confidential characteristics and file links in longitudinal linked data. In: Domingo-Ferrer, J., Torra, V. (eds.) Privacy in Statistical Databases 2004, pp. 290–297. Springer, New York, NY (2004)

Acquisti, A.: Security of personal information and privacy: technological solutions and economic incentives. In: Camp, J., Lewis, R. (eds.) The Economics of Information Security, pp. 179–186. Kluwer, Dordrecht (2004)

Adam, N.R., Wortmann, J.C.: Security-control methods for statistical databases: a comparative study. ACM Comput. Surv. **21**, 515–556 (1989)

Agrawal, R., Srikant, R.: Privacy-preserving data mining. Proceedings of the 2000 ACM SIGMOD on Management of Data, Dallas, TX, 15–18 May 2000

Bacharach, M.: Matrix rounding problem. Manage. Sci. **9**, 732–742 (1966)

Bacher, J., Brand, R., Bender, S.: Re-identifying register data by survey data using cluster analysis: an empirical study. Int. J. Uncertainty Fuzziness Knowl.-Based Syst. **10**(5), 589–608 (2002)

Backstrom, L., Dwork, C., Kleinberg, J.: Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. Proceedings of the 16th International Conference on World Wide Web, Banff, AB, 08–12 May 2007

Barabba, V.P., Kaplan, D.L.: U.S. Census Bureau Statistical Techniques to Prevent Disclosure – The Right to Privacy vs. the Need to Know. Paper presented at the 40th Session of the international Statistical Institute, Warsaw, 1975

Bayardo, R.J., Agrawal, R.: Data privacy through optimal K-anonymization. Proceedings of the 21st International Conference on Data Engineering, 2005. ICDE 2005. Tokyo (2005)

Béland, Y.: Release of public use microdata files for NPHS? mission. . .partially Accomplished! In: Proceedings of the Survey Research Methods Section, pp. 404–409. American Statistical Association, Baltimore (1999)

Benedetti, R., Franconi, L., Capobianchi, A.: Individual risk of disclosure using sampling design information. Istat Contributi n. 14/2003. Available at http://www.istat.it/dati/pubbsci/contributi/Contr_anno2003.htm (2003)

Bernardinelli, L., Montomoli, C.: Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. Stat. Med. **11**, 983–1007 (1992)

Besag, J.: Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B **36**, 192–236 (1974)

Besag, J., York, J., Mollié, A.: Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Stat. Math. **43**, 1–59 (1991)

Bethlehem, J.G., Keller, W.J., Pannekoek, J.: Disclosure control of microdata. J. Am. Stat. Assoc. **85**, 38–45 (1990)

Blakemore, M.: The potential and perils of remote access. In: Doyle, P., Lane, J., Theeuwes, J., Zayatz, L. (eds.) Confidentiality, Disclosure, and Data Access: Theory and Practical Application for Statistical Agencies, pp. 315–337. Elsevier, Amsterdam (2001)

Blien, U., Wirth, H., Müller, M.: Disclosure risk for microdata stemming from official statistics. Stat. Neerl. **46**, 69–82 (1992)

Boruch, R.F., Cecil, J.S.: Assuring the Confidentiality of Social Research Data, University of Pennsylvania Press, Philadelphia, PA 1979

Brand, R.: Microdata protection through noise addition. In: Domingo-Ferrer, J. (ed.) Inference Control in Statistical Databases. Lecture Notes in Computer Science, **vol. 2316**, pp. 97–116. Springer, Berlin, Heidelberg (2002a)

Brand, R.: Tests of the applicability of Sullivan's algorithm to synthetic data and real business data in official statistics. European Project IST-2000-25069 CASC, Deliverable 1.1-D1, http://neon.vb.cbs.nl/casc (2002b)

Brin, D.: The Transparent Society. Addison-Wesley, Reading, MA (1998)

Brownstein, J.S., Cassa, C.A., Mandl, K.D.: No place to hide – reverse identification of patients from published maps. New Engl. J. Med. 19 October **355**(16), 1741–1742 (2006)

Butz, W.P.: Data confidentiality and public perceptions: the case of the European censuses. American Statistical Association 1985 Proceedings of the Section on Survey Research Methods. American Statistical Association, Washington, DC (1985)

Butz, W.P., Scarr, H.A.: The 1987 German census: a trip report. Report prepared for the U.S. Bureau of the Census, Washington, DC, June 1987

Carter, R., Boudreau, J.-R., Briggs, M.: Analysis of the risk of disclosure for census microdata. Statistics Canada Working Paper, Statistics Canada, Ottawa, ON 1991

Carvalho, F.D., Dellaert, N.P., Osório, M.S.: Statistical disclosure in two-dimensional tables: general tables. J. Am. Stat. Assoc. **89**, 1547–1557 (1994)

Causey, B.D., Cox, L.H., Ernst, L.R.: Applications of transportation theory to statistical problems. J. Am. Stat. Assoc. **80**, 903–909 (1985)

Cecil, J.S.: Confidentiality legislation and the United States federal statistical system. J. Off. Stat. **9**(2), 5 (1993)

Chowdhury, S.D., Duncan, G.T., Krishnan, R., Roehrig, S.F., Mukherjee, S.: Disclosure detection in multivariate categorical databases: auditing confidentiality protection through two new matrix operators. Manage. Sci. **45**, 1710–1723 (1999)

Clifton, C., Kantarcioglu, M., Vaidya, J., Lin, X., Zhu, M.Y.: Tools for privacy preserving data mining. SIGKDD Explor. **4**(2), 28–34 (2002)

CNSTAT: Reengineering the survey of income and program participation committee on national statistics. Citro, C.F., Scholz, J.K. (eds.). National Academies Press, Washington, DC (2009)

Conway, R., Strip, D.: Selective partial access to a database. Proceedings of the ACM Annual Conference, New York, NY 1976

Cox, L.H.: Suppression methodology and statistical disclosure control. J. Am. Stat. Assoc. **75**, 377–385 (1980)

Cox, L.H.: Linear sensitivity measures and statistical disclosure control. J. Stat. Plann. Inference **5**, 153–164 (1981)

Cox, L.H.: Network models for complementary cell suppression. J. Am. Stat. Assoc. **90**, 1453–1462 (1995)

Cox, L.H., Kelly, J.P., Patil, R.: Balancing quality and confidentiality for multivariate tabular data. In: Domingo-Ferrer, J., Torra, V. (eds.) Privacy in Statistical Databases. Lecture Notes in Computer Science, **vol. 3050**, pp. 87–98. Springer, New York, NY (2004)

Dale, A.: Confidentiality of official statistics: an excuse for privacy. In: Dorling, D., Simpson, S. (eds.) Statistics in Society, pp. 29–37. Arnold, London (1998)

Dale, A., Elliot, M.: Proposals for 2001 SARS: an assessment of disclosure risk. J. R. Stat. Soc. Ser. A **164**(part 3), 427–447 (2001)

Dale, A., Marsh, C.: The 1991 Census Users' Guide. HMSO, London (1993)

Dalenius, T.: Finding a needle in a haystack – or identifying anonymous census records. J. Official Stat. **2**(3), 329–336 (1986)

Dalenius, T.: Controlling invasion of privacy in surveys. Department of Development and Research, Statistics, Sweden (1988)

Dalenius, T., Reiss, S.P.: Data-swapping: a technique for disclosure control (extended abstract). American Statistical Association, Proceedings of the Section on Survey Research Methods, Washington, DC, pp. 191–194 1978

Dalenius, T., Reiss, S.P.: Data-swapping: a technique for disclosure control. J. Stat. Plann. Inference **6**, 73–85 (1982)

Dandekar, R., Domingo-Ferrer, J., Sebé, F.: LHS-based hybrid microdata vs. rank swapping and microaggregation for numeric microdata protection. In: Domingo-Ferrer, J. (ed.) Inference Control in Statistical Databases. Lecture Notes in Computer Science, **vol. 2316**, pp. 153–162. Springer, Berlin, Heidelberg (2002)

De Waal, T., Willenborg, L.C.R.J.: A view on statistical disclosure for microdata. Surv. Methodol. **22**(1), 95–103 (1996)

Defays, D., Anwar, M.N.: Masking microdata using micro-aggregation. J. Official Stat. **14**, 449–461 (1998)

Defays, D., Nanopoulos, P.: Panels of enterprises and confidentiality: the small aggregates method. Proceedings of 1992 Symposium on Design and Analysis of Longitudinal Surveys, pp. 195–204. Statistics Canada, Ottawa, ON (1993)

Dellaert, N.P., Luijten, W.A.: Statistical disclosure in general three-dimensional tables. Stat. Neerl. **53**, 197–221 (1999)

DeWaal, A.G., Willenborg, L.C.R.J.: Global recodings and local suppressions in microdata sets. Proceedings of Statistics Canada Symposium' 95, pp. 121–132. Statistics Canada, Ottawa, ON (1995)

DeWaal, A.G., Willenborg, L.C.R.J.: Information loss through global recoding and local suppression. Netherlands Official Stat. **14**, 17–20 (1999). Special issue on SDC

Domingo-Ferrer, J.: On the complexity of micro aggregation. Paper presented at the UNECE Workshop on Statistical Data Editing, Skopje, Macedonia, May 2001

Domingo-Ferrer, J. (ed.): Inference Control in Statistical Databases. Springer, New York, NY (2002)

Domingo-Ferrer, J., Mateo-Sanz, J.M.: An empirical comparison of SDC methods for continuous microdata in terms of information loss and re-identification risk. Paper Presented at the UNECE Workshop on Statistical Data Editing, Skopje, Macedonia, May 2001

Domingo-Ferrer, J., Mateo-Sanz, J.M.: Practical data-oriented microaggregation for statistical disclosure control. IEEE Trans. Knowl. Data Eng. **14**(1), 189–201 (2002)

Domingo-Ferrer, J., Mateo-Sanz, J.M., Torra. V.: Comparing SDC methods for microdata on the basis of information loss and disclosure risk. Pre-proceedings of ETK-NTTS'2001, vol. 2, pp. 807–826. Eurostat, Luxemburg (2001)

Domingo-Ferrer, J., Mateo-Sanz, J.M., Oganian, A., Torres, A.: On the security of microaggregation with individual ranking: analytical attacks. Int. J. Uncertainty Fuzziness Knowl.-Based Syst. **10**(5), 477–492 (2002)

Domingo-Ferrer, J., Sebé, F., Castella, J.: On the security of noise addition for privacy in statistical databases. In: Domingo-Ferrer, J., Torra, V. (eds.) Privacy in Statistical Databases. Lecture Notes in Computer Science, **vol. 3050**, pp. 149–161. Springer, Berlin, Heidelberg (2004)

Domingo-Ferrer, J., Torra, V.: A quantitative comparison of disclosure control methods for microdata. In: Zayatz, L., Doyle, P., Theeuwes, J., Lane, J. (eds.) Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies, pp. 111–133. North-Holland, Amsterdam (2001)

Domingo-Ferrer, J., Torra, V.: Validating distance-based record linkage with probabilistic record linkage. Lecture Notes in Computer Science **2504**, 207–215 (2002)

Domingo-Ferrer, J., Torra, V.: Disclosure risk assessment in statistical microdata protection via advanced record linkage. Stat. Comput. **13**, 343–354 (2003)

Domingo-Ferrer, J., Torra, V.: Disclosure risk assessment in statistical data protection. J. Comput. Appl. Math. **164–165**, 285–293 (2004)

Domingo-Ferrer, J., Torra, V.: Ordinal, continuous and heterogeneous k-anonymity through microaggregation. Data Mining Knowl. Discov. **11**(2), 195–212 (2005)

Domingo-Ferrer, J., Torra, V. Mateo-Sanz, J.M., Sebé. F.: Research data center-based confidentiality research: systematic measures of re-identification risk based on the probabilistic links of the partially synthetic data back to the original microdata. Final Report technical report, Rovira i Virgili University and IIIACSIC, Catalonia, Spain 2005

Domingo-Ferrer, J., Torra, V. Mateo-Sanz, J.M., Sebé, F.: Empirical disclosure risk assessment of the ipso synthetic data generators. In: Santos, M.J., Bujnowska, A. (eds.) Monographs in Official Statistics-Work Session on Statistical Data Confidentiality, pp. 227–238. Eurostat, Luxemburg (2006)

Doyle, P., Lane, J., Theeuwes, J., Zayatz, L.: Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies. Elsevier Science, Amsterdam (2001)

Duncan, G.T.: Confidentiality and statistical disclosure limitation. In: Smelser, N.J., Baltes, P.B. (eds.) International Encyclopedia of the Social and Behavioral Sciences, pp. 2521–2525. Elsevier Ltd., Oxford (2001)

Duncan, G.T.: Appendix E: Confidentiality and data access issues for institutional review boards. In: Citro, C.F., Ilgen, D.R., Marrett, C.B. (eds.) Protecting Participants and Facilitating Social and Behavioral Research, pp. 235–252. National Research Council, Washington, DC (2003)

Duncan, G.T.: Exploring the tension between privacy and the social benefits of governmental databases. In: Podesta, J., Shane, P.M., Leone, R.C.A (eds.) Little Knowledge: Privacy, Security, and Public Information after September 11, pp. 71–88. The Century Foundation, New York, NY (2004)

Duncan, G.T.: Privacy by design. Science **317**, 1178–1179. August 31 (2007)

Duncan, G.T., Chowdury, S.D., Krishnan, R., Roehrig, S.F., Mukerjee, S.: Disclosure detection in multivariate categorical databases: auditing confidentiality protection through two new matrix operators. Manage. Sci. **45**, 1710–1723 (1999)

Duncan, G.T., Fienberg, S.E.: Obtaining information while preserving privacy: a Markov perturbation method for tabular data. Eurostat. Proceedings of Statistical Data Protection ′98, Lisbon, pp. 351–362 (1999)

Duncan, G.T., Fienberg, S.E., Krishnan, R., Padman, R., Roehrig, S.F.: Disclosure limitation methods and information loss for tabular data. In: Doyle, P., Lane, J., Theeuwes, J., Zayatz, L. (eds.) Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies, pp. 135–166. North-Holland, Amsterdam (2001)

Duncan, G.T., Jabine, T.B., de Wolf, V.A. (eds.): Panel on Confidentiality and Data Access, Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council and the Social Science Research Council, Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics. National Academy of Sciences, Washington, DC (1993)

Duncan, G.T., Keller-McNulty, S.A., Stokes, S.L.: Disclosure risk vs. data utility: the R-U confidentiality map. Technical report LA-UR-01-6428, Los Alamos National Laboratory, Los Alamos, NM 2001

Duncan, G.T., Lambert, D.: Disclosure-limited data dissemination (with discussion). J. Am. Stat. Assoc. **81**(393), 10–28 (1986)

Duncan, G.T., Lambert, D.: The risk of disclosure for microdata. J. Bus. Econ. Stat. **7**, 207–217 (1989)

Duncan, G.T., Mukherjee, S.: Microdata disclosure limitation in statistical databases: query size and random sample query control. Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy, Oakland, CA, pp. 278–287 (1991)

Duncan, G.T., Mukherjee, S.: Optimal disclosure limitation strategy in statistical databases: deterring tracker attacks through additive noise. J. Am. Stat. Assoc. **95**, 720–729 (2000)
Duncan, G.T., Pearson, R.W.: Enhancing access to microdata while protecting confidentiality: prospects for the future (with discussion). Stat. Sci. **6**, 219–239 (1991)
Duncan, G., Pearson, R., Jabine, T.: The pursuit of knowledge and the protection of privacy: conflicts between access to and confidentiality of surveys of U.S. doctorates. Items (Social Science Research Council), pp. 65–70 September 1989
Duncan, G.T., Roehrig, S.F.: Reconciling information privacy and information access in a globalized technology society. In: Erickson, J. (ed.) Database Technologies: Concepts, Methodologies, Tools, and Applications, pp. 1823–1843. IGI Global, Hershey, PA (2007)
Duncan, G.T., Stokes, S.L.: Disclosure risk vs. data utility: the R-U confidentiality map as applied to topcoding. Chance **17**(3), 16–20 (2004)
Dwork, C.: Ask a better question, get a better answer a new approach to private data analysis. In: Schwentik, T., Suziu, D. (eds.) 11th Proceedings of the International Conference on Database Theory, ICDT 2007, January 10–12, 2007. Lecture Notes in Computer Science, **vol. 4353**, pp. 18–27. Barcelona, Spain (2006)
Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. Proceedings of the 3rd International Conference on Very Large Data Bases, Tokyo, Japan, pp. 265–284 (2006)
Dwork, C., McSherry, F., Talwar, K.: Differentially Private Marginals Release with Mutual Consistency and Error Independent of Sample Size. Proceedings of UNECE worksession on statistical confidentiality, December 2007, pp 193–204. Manchester (2009)
Elamir, E., Skinner, C.J.: Record level measures of disclosure risk for survey microdata. J. Official Stat. **22**, 525–539 (2006)
Elliot, M.J.: DIS: a new approach to the measurement of statistical disclosure risk. Risk Manage. Int. J. **2**(4), 39–48 (2000)
Elliot, M.J.: Data intrusion simulation: advances and a vision for the future of disclosure control. Stat. J. United Nations **17**, 1–9 (2001)
Elliot, M.J.: Statistical disclosure control. In: Kempf-Leonard, K. (ed.) Encyclopedia of Social Measurement, vol. 3, pp. 663–670. Elsevier, New York, NY (2005)
Elliot, M.J.: Data Citizenship: a 21st century solution to a 20th Century problem. Keynote speech to Exploiting Existing Data for Health Research, St Andrews September (2007)
Elliot, M.J.: Privacy, Identity and Disclosure. Keynote speech to the International Conference on Communication, Computing and Security. Rourkela, February 2011. http://www.nitrkl.ac.in/conference/conference_welcome.asp?cid=30 (2011)
Elliot, M.J., Dale, A.: Disclosure risk for microdata. End of Framework IV project report to the European Union (1998)
Elliot, M.J., Dale, A.: Scenarios of attack: a data intruder's perspective on statistical disclosure risk. Netherlands Official Stat. **14**, 6–10 (1999)
Elliot, M.J., Manning, A.M., Ford, R.W.: A computational algorithm for handling the special uniques problem. Int. J. Uncertain. Fuzziness Knowl. Based Syst. **5**(10), 493–509 (2002)
Elliot, M.J., Manning, A., Mayes, K., Gurd, J., Bane, M.: SUDA: a program for detecting special uniques. Proceeding of UNECE Work Session on Statistical Data Confidentiality, Geneva, 9–11 November 2005
Elliot, M.J., Skinner, C.J., Dale, A.: Special uniques, random uniques and sticky populations: some counterintuitive effects of geographical detail on disclosure risk. Res. Official Stat. **1**(2), 53–68 (1998)
Eurostat Manual of Business Statistics: unstats.un.org/unsd/EconStatKB/Attachment252.aspx (2005)
Eurostat: 'Manual on Disclosure Control Methods'; Produced by B. Helmpecht and D. Schackis. Office for Official Publications of the European Communities, Luxembourg (1996)
Evans, T., Zayatz, L., Slanta, J.: Using noise for disclosure limitation of establishment tabular data. J. Official Stat. **14**(4), 537–551 (1998)

Evfimievski, A., Gehrke, J., Srikant, R.: Limiting privacy breaches in privacy preserving data mining. Proceedings of the 22nd ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS 2006), San Diego, CA, June 2003

Federal Committee on Statistical Methodology: Statistical Policy Working Paper 22: Report on Statistical Disclosure Limitation Methodology, U.S. Office of Management and Budget, Washington, DC 1994

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

Fellegi, I.P.: On the question of statistical confidentiality. J. Am. Stat. Assoc. **67**, 7–18 (1972)

Fellegi, I.P., Sunter, A.B.: A theory for record linkage. J. Am. Stat. Assoc. **64**, 1183–1210 (1969)

Fienberg, S.E.: Conflicts between the needs for access to statistical information and demands for confidentiality. Technical report #577, Department of Statistics, Carnegie Mellon University. Proceedings of the International Seminar on Statistical Confidentiality, International Statistical Institute, Dublin 1993

Fienberg, S.E.: Conflicts between the needs for access to statistical information and demands for confidentiality. J. Official Stat. **10**, 115–132 (1994)

Fienberg, S.E.: Confidentiality and disclosure limitation methodology: challenges for national statistics and statistical research. Commissioned by Committee on National Statistics of the National Academy of Sciences (1997)

Fienberg, S.E.: Confidentiality and disclosure limitation. In: Kempf-Leonard, K. (ed.) Encyclopedia of Social Measurement, pp. 463–469. Elsevier, New York, NY (2005)

Fienberg, S.E., McIntyre, J.: Data swapping: variations on a theme by Dalenius and Reiss. In: Domingo-Ferrer, J., Torra, V. (eds.) PSD 2004. Lecture Notes in Computer Science, **vol. 3050**, pp. 14–29. Springer, Berlin, Heidelberg (2004)

Fienberg, S.E., Makov, U.E., Sanil, A.P.: A Bayesian approach to data disclosure: optimal intruder behavior for continuous data. J. Official Stat. **14**, 75–89 (1997)

Fienberg, S.E., Makov, U.E., Steel, R.J.: Disclosure limitation using perturbation and related methods for categorical data. J. Official Stat. **14**, 485–502 (1998)

Fienberg, S.E., Steele, R.J., Makov, U.E.: Statistical notions of data disclosure avoidance and their relationship to traditional statistical methodology: data swapping and loglinear models. Proceedings of Bureau of the Census 1996 Annual Research Conference, US Bureau of the Census, Washington, DC, pp. 87–105 1996

Fisch, K., McLeod, S.: Did you know 2.0 US Chamber of Commerce, Washington, DC, June (2007)

Fischetti, M., Salazar, J.J.: Computational experience with the controlled rounding problem in statistical disclosure control. J. Official Stat. **14**(4), 553–565 (1998)

Fischetti, M., Salazar, J.J.: Models and algorithms for the 2-dimensional cell suppression problem in statistical disclosure control. Math. Program. **84**, 283–312 (1999)

Fischetti, M., Salazar, J.J.: Models and algorithms for optimizing cell suppression in tabular data with linear constraints. J. Am. Stat. Assoc. **95**(451), 916–928 (2000)

Fischetti, M., Salazar, J.J.: Partial cell suppression: a new methodology for statistical disclosure control. Stat. Comput. **13**, 13–21 (2003)

Foster, L., Haltiwanger, J., Krizan, C.J.: Aggregate productivity growth: lessons from microeconomic evidence. In: Hulten, C.R., Dean, E.R., Harper, M.J. (eds.) New Developments in Productivity Analysis, pp. 303–363. University of Chicago Press, Chicago, IL (2001)

Franconi, L., Stander, J.: Model based disclosure limitation for business microdata. The Statistician **51**, 51–61 (2002)

Franconi, L., Stander, J.: Spatial and non-spatial model-based protection for the release of business microdata. Stat. Comput. **13**, 295–305 (2003)

Freeman, L.: The Development of Social Network Analysis. Empirical Press, Vancouver, BC (2006)

Fuller, W.A.: Masking procedures for microdata disclosure limitation. J. Official Stat. **9**, 383–406 (1993)
Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman, New York, NY. ISBN 0-7167-1045-5 (1979)
Garfinkel, R., Gopal, R., Goes, P.: Privacy protection of binary confidential data against deterministic, stochastic, and insider attack. Manage. Sci. **48**, 749–764 (2002)
Gehrke, J.: Models and methods for privacy-preserving data publishing and analysis (tutorial slides). Twelfth Annual SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD 2006), Philadelphia 2006
Geurts, J.: Heuristics for cell suppression in tables. Master's thesis, Netherlands Central Bureau of Statistics, Voorburg (1992)
Gill, L.: Methods for automatic record matching and linking and their use in national statistics. National Statistics Methodology Series, no. 25. Office for National Statistics, London (2001)
Goldstein, H.: Multilevel Models in Educational and Social Research. Charles Griffin, London (1987)
Gomatam, S., Larsen, M.D.: Record linkage and counterterrorism. Chance **17**(1), 25–29 (2004)
Gomatam, S., Karr, A.F., Sanil, A.P.: Data Swapping as a Decision Problem. J. Off. Stat. **21**(4), 635–655 (2005)
Greely H., Collecting biomeasures in the panel study of income dynamics: ethical and legal concerns. Biodemography Soc. Biol. **55**(2), 270–288 (2009)
Greenberg, B.V.: Disclosure avoidance research at the census Bureau. Proceedings of the Bureau of the Census Sixth Annual Research Conference, Bureau of the Census, Washington, DC, pp. 144–166 1990
Greenburg, B.V., Voshell, L.: The geographic component of disclosure risk for microdata. SRD Research report Census/SRD/RR-90/13, Bureau of the Census, Washington, DC 1990
Griffin, R.A., Navarro, A., Flores-Baez, L.: Disclosure avoidance for the 1990 Census. Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA, pp. 516–521 (1989)
Hansen, S.L., Mukherjee, S.: A polynomial algorithm for optimal univariate microaggregation. IEEE Trans. Knowl. Data Eng. **15**(4), 1043–1044 (2003)
Heitzig, J.: The Jackknife method: confidentiality protection for complex statistical analyses. Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality. Geneva, Switzerland, 9–11 November (2005)
Huizinga, D., et al.: Childhood maltreatment, subsequent antisocial behavior, and the role of monoamine oxidase A genotype. Biol. Psychiatry **60**(7), 677–683 1 Oct (2006)
Jabine, T.B.: Procedures for restricted data access. J. Official Stat. **9**(2), 537–589 (1993a)
Jabine, T.B.: Statistical disclosure limitation practices of United States statistical agencies. J. Official Stat. **9**(2), 427–454 (1993b)
Jaro, M.A.: Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. J. Am. Stat. Assoc. **84**(406), 414–420 (1989)
Jewett, R.: Disclosure analysis for the 1992 economic census. Technical report, U.S. Bureau of the Census, Washington, DC 1993
Kamlet, M.S., Klepper, S., Frank, R.G.: Mixing micro and macro data: statistical issues and implication for data collection and reporting. Proceedings of the 1983 Public Health Conference on Records and Statistics, U.S. Department of Health and Human Services, Hyattsville, MD 1985
Keller, W.J., Bethlehem, J.G.: Disclosure protection of microdata: problems and solutions. Stat. Neerl. **46**, 5–19 (1992)
Kelly, J.P.: Confidentiality protection in two and three-dimensional tables. Ph. D. thesis, University of Maryland, College Park, MD (1990)
Kelly, J.P., Golden, B.L., Assad, A.A.: Using simulated annealing to solve controlled rounding problems. ORSA J. Comput. **2**, 174–185 (1990)

Kelly, J.P., Golden, B.L., Assad, A.A.: Cell suppression: disclosure protection for sensitive tabular data. Networks **22**, 397–417 (1992)
Kelly, J.P., Golden, B.L., Assad, A.A.: Large-scale controlled rounding using tabu search with strategic oscillation. Ann. Oper. Res. **41**, 69–84 (1993)
Kennickell, A.B.: Multiple imputation and disclosure control: the case of the 1995 survey of consumer finances. In: Alvey, W., Jamerson, B., (eds.) Record Linkage Techniques 1997, pp. 248–267. National Academy Press, Washington, DC. http://www.fcsm.gov (1999)
Kennickell, A.B., Lane, J.: Measuring the impact of data protection techniques on data utility: evidence from the survey of consumer finances. In: Domingo-Ferrer, J. (ed.) Privacy in Statistical Databases. Lecture Notes in Computer Science, **vol. 4302**, pp. 291–303. Springer, New York, NY (2006)
Kleinberg, J.M.: Challenges in mining social network data: processes, privacy, and paradoxes. Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, 12–15 August, pp. 4–5 2007
Kim, J.J.: A method for limiting disclosure in microdata based on random noise and transformation. Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA, pp. 370–374 1986
Kim, J.J., Winkler, W.E.: Masking microdata files. Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA, pp. 114–119 (1995)
Kim, J.J., Winkler, W.E.: Multiplicative noise for masking continuous data. Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA 2001
Kirkwood, C.W.: Strategic Decision Making: Multi objective Decision Analysis with Spreadsheets. Duxbury Press, Belmont, CA, (1996)
Kooiman, P., Nobel, J., Willenborg, L.: Statistical data protection at statistics Netherlands. Netherlands Official Stat. **14**, 21–25 (1999)
Kirkwood, C.W.: Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets. Duxbury Press, Belmont, CA (1996)
Lambert, D.: Measures of disclosure risk and harm. J. Official Stat. **9**, 313–331 (1993)
Lane, J.: Key Issues in Confidentiality Research: Results of an NSF Workshop. http://www.nsf.gov/sbe/ses/mms/nsfworkshop_summary1.pdf (2003)
Laszlo, M., Mukherjee, S.: Minimum spanning tree partitioning algorithm for microaggregation. IEEE Trans. Knowl. Data Eng. **17**(7), 902–911 (2005)
Lee, J., McClellan, M.B., Skinner, J.S.: The distributional effects of medicare (January 1999). NBER Working Paper No. W6910 (1999)
Little, R.J.A.: Statistical analysis of masked data. J. Official Stat. **9**(2), 407–426 (1993)
Little, R.J.A., Liu, F.: Selective multiple imputation of keys for statistical disclosure-control in microdata. Proceedings of the Section on Survey Research Methods, CD-ROM, American Statistical Association, Alexandria, VA 2002
Little, R.J.A., Liu, F.: Comparison of SMIKe with data-swapping and PRAM for statistical disclosure control of simulated microdata. Proceedings of the Section on Survey Research Methods, American Statistical Association, Seattle (2003)
Loeve, A.: Notes on sensitivity measures and protection levels. Technical report, Statistics Netherlands, The Hague (2001)
Machanavajjhala, A., Kifer, D., JAbowd, J., Gehrke, J., Vilhuber, L.: Privacy: from theory to practice on the map. ICDE Conference 2008. Cancun, Mexico, April 2008. http://www.cs.cornell.edu/johannes/papers/2008/icde2008-privacy.pdf (2008)
Mackey, E.: A framework for understanding statistical disclosure processes: a case study using the UK's neighbourhood statistics. PhD Thesis, Manchester University (2009)
Mackey, E., Elliot, M.J.: An application of game theory to understanding disclosure events. Proceedings of Work Session on Statistical Data Confidentiality, Bilbao, December 2009
Mackie, C., Bradburn, N.: Improving Access to and Confidentiality of Research Data. National Academy Press, Washington, DC (2000)
Marsh, C., Skinner, C., Arber, S., Penhale, B., Openshaw, S., Hobcraft, J., Lievesley, D., Walford, N.: The case for samples of anonymized records from the 1991 census. J. R. Stat. Soc. Ser. A **154**, 305–340 (1991)

Martin, D., Kifer, D., Machanavajjhala, A., Gehrke, J., Halpern, J.: Worst case background knowledge for privacy preserving data publishing. In Proceedings of the 23rd International conference on Data Engineering, ICDE 2007, April 15–20, Istanbul, Turkey (2007)

Mateo-Sanz, J.M., Domingo-Ferrer, J.: A method for data-oriented multivariate microaggregation. In: Domingo-Ferrer, J. (ed.) Statistical Data Protection, pp. 89–99. Office for Official Publications of the European Communities, Luxemburg (1999)

Mateo-Sanz, J.M., Sebé, F., Domingo-Ferrer, J.: Outlier protection in continuous microdata masking. In: Domingo-Ferrer, J., Torra, V. (eds.) Privacy in Statistical Databases. Lecture Notes in Computer Science, **vol. 3050**, pp. 201–215. Springer, Berlin, Heidelberg (2004)

McCullagh, K.: Data sensitivity: proposals for resolving the conundrum. J. Int. Commer. Law Technol. **2**(4), 190–201 (2007)

McCaa, R., Odinga, A.: Statistical confidentiality and the construction of anonymized public use census samples: a draft proposal for the Kenyan Microdata for 1989. Paper presented to Social Science History Annual Convention, Chicago, IL, 15–18 November 2001

McCaa, R., Ruggles, S.: The census in global perspective and the coming microdata revolution. Scand. Populat. Stud. **17**, 7–30 (2002)

McDonald, S.-K.: Recent experiences with resistance to national censuses in Western Europe. U.S. Bureau of the Census memorandum from Sarah-Kathryn McDonald, October 17, (1984)

McLuhan, M.: Understanding Media: The Extensions of Man. McGraw-Hill, New York, NY (1964)

McMillen, M.: Data access: national center for education statistics, of significance. J. Assoc. Public Data Users **2**, 1 (2001)

Mera, R.N.: Matrix masking methods which preserve moments. American Statistical Association Proceedings http://www.amstat.org/sections/srms/proceedings/papers/1997_075.pdf. Anaheim (1997)

Mokken, R.J., Kooiman, P., Pannekoek, J., Willenborg, L.C.R.J.: Disclosure risks for microdata. Stat. Neerl. **46**, 49–67 (1992)

Mollié, A.: Bayesian mapping of disease. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (eds.) Markov Chain Monte Carlo in Practice, pp. 359–379. Chapman and Hall, London (1996)

Moore, R.: Controlled data swapping techniques for masking public use data sets. U.S. Bureau of the Census, Statistical Research Division Report rr96/04. Available at http://www.census.gov/srd/papers/pdf/rr96-4.pdf. Accessed Jan 21, 2011 (1996)

Müller, W., Blien, U., Wirth, H.: Identification risks of micro data. Evidence from experimental studies. Sociol. Methods Res. **24**, 131–157 (1995)

Muralidhar, K., Parsa, R., Sarathy, R.: A general additive data perturbation method for database security. Manage. Sci. **45**(10), 1399–1415 (1999)

National Research Council: Expanding access to research data: reconciling risks and opportunities. Panel on Data Access for Research Purposes, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. The National Academies Press, Washington, DC (2005)

National Research Council: Putting people on the map: protecting confidentiality with linked social-spatial data. Panel on confidentiality issues arising from the integration of remotely sensed and self-identifying data. In: Gutmann, M.P., Stern, P.C. (eds.) Committee on the Human Dimensions of Global Change, Division of Behavioral and Social Sciences and Education. The National Academies Press, Washington, DC. http://books.nap.edu/catalog.php?record_id=11865 (2007)

Navarro, A., Flores-Baez, L., Thompson, J.: Results of data switching simulation. Presented at the Spring meeting of the American Statistical Association and Population Statistics Census Advisory Committees, Washington, DC (1988)

National Academy Press: Expanding Access to Research Data: Reconciling Risks and Opportunities. http://www.nap.edu/openbook.php?record_id=11434&page=66 (2005)

Oganian, A., Domingo-Ferrer, J.: On the complexity of optimal microaggregation for statistical disclosure control. Stat. J. United Nations Econ. Commis. Eur. **18**, 4:345–354 (2001)

O'Keefe, C., Good, N.M.: A remote analysis server – what does regression output look like? In: Domingo-Ferrer, J., Saygýn, Y. (eds.) Privacy in Statistical Databases. Lecture Notes in Computer Science, **vol. 5262**, pp. 270–283. Springer, Berlin (2008)

Paass, G.: Disclosure risk and disclosure avoidance for microdata. J. Bus. Econ. Stat. **6**(4), 487–500 (1988)

Pagliuca, D., Seri, G.: Some Results of Individual Ranking Method on the System of Enterprise Accounts Annual Survey, Esprit SDC Project, Deliverable MI-3/D2 (1999)

Pickle, L.W., Waller, L.A., Lawson, A.B.: Current practices in cancer spatial data analysis: a call for guidance. Int. J. Health Geogr. **4**, 3 (2005)

Polettini, S.: Maximum entropy simulation for microdata protection. Stat. Comput. **13**(4), 307–320 (2003)

Polettini, S., Stander, J.: A Bayesian hierarchical model approach to risk estimation in statistical disclosure limitation. In Domingo-Ferrer, J., Torra, V. (eds.) Privacy in Statistical Databases, pp. 247–261. Springer, Berlin (2004)

Purdam, K., Elliot, M.J.: An evaluation of the availability of public data sources which could be used for identification purposes – A Europe wide perspective, CASC project. University of Manchester, Manchester (2002)

Purdam, K., Elliot, M.J.: A case study of the impact of statistical disclosure control on data quality in the individual UK samples of anonymised records. Environ. Plann. A **39**, 1101–1118 (2007)

Purdam, K., Mackey, E., Elliot, M.J.: Whose data is it? Personal data and privacy. Paper presented to British Sociology Association Annual Conference, York 2003a

Purdam, K., Mackey, E., Elliot, M.J.: Personal data, privacy and the 2001 UK census. Paper presented to International Conference for the Development of the Information Society Conference, Lisbon, Portugal 2003b

Raghunathan, T.E.: Evaluation of inferences from multiple synthetic data sets created using semiparametric approach. Panel on Confidential Data Access for Research Purposes, Committee on National Statistics, October (2003)

Raghunathan, T.E., Reiter, J.P., Rubin, D.R.: Multiple imputation for statistical disclosure limitation. J. Official Stat. **19**, 1–16 (2003)

Raghunathan, T.E., Rubin, D.R.: Multiple imputation for disclosure limitation. Department of Biostatistics Technical Report, University of Michigan, Dearborn, MI 2000

Rasinski, K.A., Wright, D.: Practical aspects of disclosure analysis, of significance: a. Topical J. Assoc. Public Data Users **2**(1), 35–41 (2000)

Rastogi, V., Suciu, D., Hong, S.: The boundary between privacy and utility in data publishing. Technical Report, University of Washington, Washington, DC 2007

Reiss, S.P.: Practical data-swapping: the first steps. ACM Trans. Database Syst. **9**, 20–37 (1984)

Reiss, S.P., Post, M.J., Dalenius, T.: Non-reversible privacy transformations. Proceedings of the ACM Symposium on Principles of Database Systems, Los Angeles, pp. 139–146 1982

Reiter, J.P.: Satisfying disclosure restrictions with synthetic data sets. J. Official Stat. **18**(4), 531–544 (2002)

Reiter, J.P.: Inference for partially synthetic, public use microdata sets. Surv. Methodol. **29**, 181–188 (2003)

Reiter, J.P.: Releasing multiply-imputed, synthetic public use microdata: an illustration and empirical study. J. R. Stat. Soc. Ser. A **168**, 185–205 (2005a)

Reiter, J.P.: Estimating risks of identification disclosure for microdata. J. Am. Stat. Assoc. **100**, 1103–1113 (2005b)

Reiter, J.P.: Releasing multiply imputed, synthetic public-use microdata: an illustration and empirical study. J. R. Stat. Soc. A **168**, 185–205 (2005c)

Reiter, J.P.: Significance tests for multi-component estimands from multiply-imputed, synthetic microdata. J. Stat. Plann. Inference **131**, 365–377 (2005d)

"Restricted Access Procedures" by the Confidentiality and Data Access Committee (April 2002) at http://www.fcsm.gov/committees/cdac/cdacra9.doc

Robert, C.P., Casella, G.: Monte Carlo Statistical Methods (second edition). Springer, New York, NY (2004)

Robertson, D.A.: Cell suppression at statistics Canada. Proceedings of the Second International Conference on Statistical Confidentiality, Luxembourg 1994

Robertson, D.A.: Improving Statistics Canada's cell suppression software. Technical report, Statistics Canada, Ottawa, ON 2000

Robertson, D.A., Ethier, R.: Cell suppression: experience and theory. In: Domingo-Ferrer, J. (ed.) Inference Control in Statistical Databases: From Theory to Practice. Lecture Notes in Computer Science, **vol. 2316**, pp. 8–20. Springer, New York, NY (2002)

Rubin, D.B.: Discussion of statistical disclosure limitation. J. Official Stat. **9**(2), 461–468 (1993)

Rubin, D.B.: Satisfying confidentiality constraints through the use of synthetic multiply-imputed microdata. J. Official Stat. **9**, 461–468 (1996)

Ruggles, S.: The public use microdata samples of the U.S. Census: research applications and privacy issues. A report of the Task Force on Census 2000, Minnesota Population Center and Inter-University Consortium for Political and Social Research Census 2000 Advisory Committee, Minnesota (2000)

Salazar, J.J.: A unified mathematical programming framework for different statistical disclosure limitation methods. Oper. Res. **53**(3), 819–829 (2005)

Salazar, J.J.: Controlled rounding and cell perturbation: statistical disclosure limitation methods for tabular data. Math. Program. **105**(2–3), 251–274 (2006a)

Salazar, J.J.: A new approach to round tabular data. In: Domingo-Ferrer, J., Franconi L. (eds.) Privacy in Statistical Databases. Lecture Notes in Computer Science, **vol. 4302**, pp. 25–34. Springer, Berlin, Heidelberg (2006b)

Salazar, J.J.: Statistical confidentiality: optimization techniques to protect tables. Comput. Oper. Res. **35**, 1638–1651 (2008)

Salazar, J.J.: Branch-and-cut versus cut-and-branch algorithms for cell suppression. In: Domingo-Ferrer, J., Magkos, E. (eds.) Privacy in Statistical Databases. Lecture Notes in Computer Science, **vol. 6344**, pp. 29–40. Springer, Berlin, Heidelberg (2010)

Salazar, J.J., Lowthian, P., Young, C., Merola, G., Bond, S., Brown, D.: Getting the best results in controlled rounding with the least effort. In: Domingo-Ferrer, J. (ed.) Privacy in Statistical Databases. Lecture Notes in Computer Science, **vol. 3050**, pp. 58–72. Springer, New York, NY (2004)

Samuels, S.J.: A Bayesian species-sampling-inspired approach to the uniques problem in microdata disclosure risk assessment. J. Official Stat. **14**, 373–384 (1998)

Sande, G.: Automated cell suppression to preserve confidentiality of business statistics. Stat. J. United Nat. ECE **2**, 33–41 (1984)

Sande, G.: Structure of the ACS automated cell suppression system, In Statistical Data Confidentiality. Proceedings of the Joint Eurostat/UN-ECE Work Session on Statistical Confidentiality, Skopje, pp. 105–121 (1999)

Sande, G.: Exact and approximate methods for data directed microaggregation in one or more dimensions. Int. J. Uncertain. Fuzziness Knowl. Based Syst. **10**(5), 459–476 (2002)

Sanil, A., Gomatam, S., Karr, A.: NISS WebSwap: a web-service for data swapping. J. Stat. Softw. 8(7) (2003). http://www.jstatsoft.org/v08/i07

Schlörer, J.: Security of statistical databases: multidimensional transformation. ACM Trans. Database Syst. **6**(1), 95–112 (1981)

Shlomo, N.: Accessing microdata via the internet. Joint UN/ECE and Eurostat Work Session on Statistical Data Confidentiality, Working Paper No. 6. Luxembourg, April 7–9 (2003)

Sieber, J.E.: Privacy and Confidentiality: As Related to Human Research in Social and Behavioral Science (Research Involving Human Participants V2): Online Ethics Center for Engineering, May 25, National Academy of Engineering www.onlineethics.org/CMS/research/resref/nbacindex/nbachindex/hsieber.aspx. Accessed Sept 25, 2007 (2007)

Simard, M.: Development of a real time remote access infrastructure at statistics Canada. Paper presented at the joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Bilbao, Spain, 2–4 December 2009

Singer, E., Mathiowetz, N.A., Couper, M.P.: The impact of privacy and confidentiality concerns on survey participation: the case of the 1990 U.S Census. Public Opin. Q. **57**(4), (Winter 1993), 465–482 (1993)

Singer, E., Van Hoewyk, J., Neugebauer, R.J.: Attitudes and behavior – the impact of privacy and confidentiality concerns on participation in the 2000 census. Public Opin. Q. **67**, 368–384 (2003)

Skinner, C.J.: Statistical disclosure issues for census microdata. Paper presented at International Symposium on Statistical Disclosure Avoidance, Voorburg, The Netherlands, 13 December 1990

Skinner, C.J., Elliot, M.J.: A measure of disclosure risk for microdata. J. R. Stat. Soc. Ser. B **64**(4), 855–867 (2002)

Skinner, C.J., Holmes, D.J.: Modelling population uniqueness. Proceedings of the International Seminar on Statistical Confidentiality, International Statistical Institute, Luxembourg, pp. 175–199 (1992)

Skinner, C.J., Shlomo, N.: Assessing identification risk in survey micro-data. J. Am. Stat. Assoc. **103**(483), 989–1001 (2008)

Skinner, C.J., Holmes, D.J.: Estimating the re-identification risk per record in microdata. J. Off. Stat. **14**, 361–372 (1998)

Sparks, R., Carter, C., Donnelly, J., O'Keefe, C.M., Duncan, J., Keighley, T., McAullay, D.: Remote access methods for exploratory data analysis and statistical modelling: privacy-preserving analytics TM. Comput. Methods Progr. Biomed. Arch. **91**(3), 208–222 (2008)

Smith, D., Elliot, M.J.: A measure of disclosure risk for aggregate data. Paper presented to the International Conference of the Royal Statistics Society, Turin, September 2004

Smith, D., Elliot, M.J.: An experiment in Naive Bayesian record linkage. Proceedings of Conference of the International Statistical Institute, Sydney, April 2005

Smith, D., Elliot, M.J.: A measure of disclosure risk for tables of counts. Trans. Data Priv. **1**(1), 34–52 With Smith, D. (2008)

Spruill, N.L.: The confidentiality and analytic usefulness of masked business microdata. Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 602–607. Alexandria, VA (1983)

Statistics Netherlands: Argus User's Manual. http://neon.vb.cbs.nl/casc/ (2007)

Steel, P., Sperling, J.: The Impact of Multiple Geographies and Geographic Detail on Disclosure Risk: Interactions between Census Tract and ZIP Code Tabulation Geography. U.S. Census Bureau, Washington (2001)

Stigler, S.M.: Adolphe Quetelet. Encyclopedia of Statistical Sciences. Wiley, New York, NY (1986)

Sullivan, G., Fuller, W.A.: The use of measurement error to avoid disclosure. Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA, pp. 802–807 (1989)

Sweeney, L.: Achieving k-anonymity privacy protection using generalization and suppression. Int. J. Uncertain. Fuzziness Knowl. Based Syst. **10**(5), 571–588 (2002)

Takemura, A.: Local recoding and record swapping by maximum weight matching for disclosure control of microdata sets. J. Official Stat. **18**(2), 275–289 (2002)

Tendick, P.: Optimal noise addition for preserving confidentiality in multivariate data. J. Stat. Plann. Inference **27**, 341–353 (1991)

Tendick, P., Matloff, N.: A modified random perturbation method for database security. ACM Trans. Database Syst. **19**, 47–63 (1994)

Thibaudeau, Y., Winkler, W.E.: Bayesian networks representations, generalized imputation, and synthetic microdata satisfying analytic restraints. Statistical Research Division Report RR 2002/09, Washington. http://www.census.gov/srd/www/byyear.html (2002)

Torra, V.: Microaggregation for categorical variables: a median based approach. In: Domingo-Ferrer, J., Torra, V. (eds.) Privacy in Statistical Databases. Lecture Notes in Computer Science, **vol. 3050**, pp. 162–174. Springer, Berlin, Heidelberg (2004)

Torra, V., Abowd, J., Domingo-Ferrer, J.: Using Mahalanobis distance-based record linkage for disclosure risk assessment. Lect. Notes in Comput. Sci. **4302**, 233–242 (2006)

Tranmer, M., Pickles, A., Fieldhouse, E., Elliot, M.J., Dale, A., Brown, M., Martin, D., Steel, D., Gardiner, C.: Microdata for small areas. J. R. Stat. Soc. Ser. A **168**(1), 29–49 (2005)

Tranmer, M., Steel, D., Chambers, R., Clark, R., Elliot, M.: The role of individuals, geographical groups, households and social networks in social statistics. Paper presented to the 30th Sunbelt Conference Riva del Garda, Italy, June 2010

Trottini, M.: A decision-theoretic approach to data disclosure problems. Paper prepared for 2nd Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Skopje, Macedonia, 14–16 March 2001 (2001)

Trottini, M.: Decision models for data disclosure limitation. Ph.D. thesis, Department of Statistics, Carnegie Mellon University (2003)

Trottini, M.: Statistical disclosure limitation in longitudinal linked data: objectives and attributes. UNECE/Eurostat Work Session on Statistical Data Confidentiality, Geneva, Monograph in Official Statistics, Eurostat, 165–173 (2005)

United Nations: Managing Statistical Confidentiality & Microdata Access: Principles and Guidelines of Good Practice (2007)

Vaidya, J., Clifton, C., Zhu, M.: Privacy Preserving Data Mining. Springer, Heidelberg (2006)

Valls, V., Torra, V., Domingo-Ferrer, J.: Aggregation methods to evaluate multiple protected versions of the same confidential data set. In: Grzegorzewski, P., Hryniewicz, O., Gil, M.A. (eds.) Soft Methods in Probability, Statistics and Data Analysis, pp. 355–362. Series Advances in Soft Computing. Physica, Heidelberg (2002)

Willenborg, L.C.R., de Waal, T.: Statistical Disclosure Control in Practice. Lecture Notes in Statistics, **vol. 111**. Springer, New York, NY (1996)

Willenborg, L.C.R., de Waal, T.: Elements of Statistical Disclosure Control. Lecture Notes in Statistics, **vol. 155**. Springer, New York, NY (2001)

Winkler, W.E.: Matching and record linkage. Technical report RR93/08, Statistical Research Division, U.S. Bureau of the Census, Washington, DC 1993

Winkler, W.E.: Advanced methods for record linkage. Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, VA, pp. 467–472 1994

Winkler, W.E.: Matching and record linkage. In: Cox, B.G. et al. (ed.) Business Survey Methods, pp. 355–384. Wiley, New York, NY (1995a)

Winkler, W.E.: Advanced methods for record linkage. Proceedings of the American Statistical Association Section on Survey Research Methods, Alexandria, VA, pp. 467–472 1995b

Winkler, W.E.: Re-identification methods for evaluating the confidentiality of analytically valid microdata. Res. Official Stat. **1**, 87–104 (1998)

Winkler, W.E.: Masking and re-identification methods for public-use microdata: overview and research problems. RESEARCH REPORT SERIES (Statistics #2004-06): Statistical Research Division, U.S. Bureau of the Census http://www.census.gov/srd/papers/pdf/rrs2004-06.pdf (2004). Accessed Oct 5, 2007

Winkler, W.E.: Masking and re-identification methods for public-use microdata: overview and research problems. In: Domingo-Ferrer, J., Torra, V. (eds.) Privacy in Statistical Databases. Lecture Notes in Computer Science, **vol. 3050**, pp. 231–246. Springer, Berlin, Heidelberg. http://www.census.gov/srd/papers/pdf/rrs2004-06.pdf (2004a)

Winkler, W.E.: Re-identification methods for masked microdata. In: Domingo-Ferrer, J., Torra, V. (eds.) Privacy in Statistical Databases, pp. 216–230. Springer, New York, NY (2004b)

Wu, J., Abowd, J.: Synthetic data for administrative record applications at LEHD. Available online in the LEHD Presentations Library (2008)

Xiao, X., Wang, G., Gehrke, J.: Interactive anonymization of sensitive data. SIGMOD'09 June 29–July 2, 2009, Providence, Rhode Island, USA. ACM 978-1-60558-551-2/09/06 (2009)

Yancey, W.E., Winkler, W.E., Creecy, R.H.: Disclosure risk assessment in perturbative microdata protection. In: Domingo-Ferrer, J. (ed.) Inference Control in Statistical Databases. Lecture Notes in Computer Science, **vol. 2316**, pp. 135–152. Springer, Berlin, Heidelberg (2002)

Zaslavsky, A.M., Horton, N.J.: Balancing disclosure risk against the loss of nonpublication. J. Official Stat. **14**, 411–419 (1998)

Zayatz, L.V.: Estimation of the percent of unique population elements in a microdata file using the sample. Statistical Research Division Report Series, Census/SRD/RR-91/08 Bureau of the Census, Washington, DC, pp. 674–684 1991

Zayatz, L.V.: Using linear programming methodology for disclosure avoidance purposes, Technical report, U.S. Bureau of the Census. Research Report RR-92/02, Washington, DC 1992

Zayatz, L.V., Massell, P., Steel, P.: Disclosure limitation practices and research at the U.S. Census Bureau. Special issue on statistical disclosure control. Neth. Off. Stat. **14**, 26–29 (1999)

Zayatz, L.V., Rowland, S.: Disclosure limitation for American FactFinder. Paper presented at the American Statistical Association Joint Statistical Meetings, Baltimore, MD, 8 August 1999

Index