製表関連国際用語集 No.2

統計データ機密保護に関する
国連欧州経済委員会／EU 統計局
合同ワークセッション

統計データ開示抑制に関する用語集
改訂版（対訳）

2005 年 8 月

（翻訳：独立行政法人 統計センター　研究センター，平成 18 年６月）

統計データ開示抑制に関する用語集改訂版（対訳）

Mark　Elliot（マンチェスター大学）
Anco　Hundepool（オランダ統計局）
Eric　Schulte　Nordholt（オランダ統計局）
Jean-Louis　Tambay（カナダ統計局）
Thomas　Wende（ドイツ統計局）

2005 年８月版

序文

本用語集の構想は、ルクセンブルグで開催された統計データ機密保護に関するECE/Eurostat 合同ワークセッション（Joint ECE/Eurostat Work Session on Statistical Data Editing, 2003 年 4 月 7−9 日）で打ち出された。本用語集の作成にあたった 5 名は、すべて同ワークセッションの出席者であり、2003 年 8 月 18 日にベルリンで開かれた ISI セッションにも参加した人々である。この統計開示抑制に関する新しい用語集は、ジュネーブで開催の統計データ機密保持に関するECE/Eurostat次回合同ワークセッション（2005 年 11 月 9−11 日）で発表される予定である。そのために、世界中の専門家からコメントをいただけるよう準備版を発表した。本用語集には 2 つの目的がある。第一は、この分野に詳しくない人に統計データ開示管理で使用される用語をよく知ってもらうことであり、第二は、同時に統計データ開示管理についての研修の中で参考書として使用することもできる。我々は、本用語集が役立ち、その 2 つの目的が達せられることを望むものである。ご意見あるいはお問合せがある場合には、今後の版に反映できるよう Eric Schulte Nordholt（e-mail：ELSE@CBS.NL）まで連絡をいただきたい。

謝辞

この資料（草案）作成における多くの方々のご協力に感謝する。特に、Paul Feuvrier(INSEE)、Kingsley Purdam(マンチェスター大学)、Barry Schouten(オランダ統計局)、Duncan Smith(マンチェスター大学)及び Peter-Paul de Wolf(オランダ統計局)。

訳注．用語の中には定訳がない，あるいは，対応する適切な日本語が存在しないか標準的なものがないと思われるものが多く，本資料で用いた訳語の中には適訳とは言い難いものが含まれている恐れがある。このため，対訳形式を採り，適宜，原文を参照できるようにした。また、参考のため、本改訂版作成に当たり、和訳の見出し語を変更したものについて、巻末に一覧にし、EU の CASC プロジェクト（Computational Aspects of Statistical Disclosure）のホームページに掲載された 2005 年 9 月版に合わせ、一部変更を行った。

A	A
Ambiguity rule: Synonym of (p,q) rule.	**曖昧性ルール：**（p,q）ルールと同義
Analysis server: A form of **remote data** laboratory designed to run analysis on data stored on a safe server. The user sees the results of their analysis but not the data.	**分析サーバー：**安全なサーバーに保存されたデータの分析を実行するように設計されたリモート・データ・ラボラトリーの一形態。ユーザーは、データではなくその分析結果を参照する。
Anonymised data: Data containing only anonymised records.	**匿名化データ：**匿名化レコードだけを含んだデータ
Anonymised records: A record from which direct identifiers have been removed.	**匿名化レコード：**直接識別子を取り除いたレコード
Approximate disclosure: Approximate disclosure happens if a user is able to determine an estimate of a respondent value that is close to the real value. If the estimator is exactly the real value the disclosure is exact.	**近似的開示：**近似的開示は、ユーザーが回答者の数値の実際値に近い推定値を決定することができる場合に発生する。推定値が実際の値と完全に一致する場合には、完全な開示になる。
Argus: Two software packages for Statistical Disclosure Control are called Argus. μ-Argus is a specialized software tool for the protection of **microdata**. The two main techniques used for this are **global recording** and **local suppression**. In the case of global recoding several categories of a variable are collapsed into a single one. The effect of local suppression is that one or more values in an unsafe combination are suppressed, i.e. replaced by a missing value. Both **global recoding** and **local suppression** lead to a loss of information, because either less detailed information is provided or some information is not given at all. τ-Argus is a specialized software tool for the protection of tabular data. *τ-Argus is used to produce safe tables.* τ-Argus uses the same two main techniques as μ-Argus: global recoding and local suppression. For τ-Argus the latter consists of suppression of cells in a table.	**ARGUS：**統計データ開示抑制を目的とした２つのソフトウェア・パッケージである。μ-Argus は、ミクロデータ保護のためのソフトウェアツールである。このために利用される 2 つの主要技術は、大域的再符号化と局所秘匿である。大域的再符号化の場合には、変数の複数のカテゴリーを一つに統合する。局所秘匿は、安全でない一つの値あるいは複数の値の組み合わせを秘匿する、すなわち、欠損値に置き換えることである。大域的再符号化と局所秘匿はいずれも情報ロスを招く。これは、情報の詳細度が低下したり、一部の情報が全く提供されないためである。τ-Argus は、集計表データ保護を目的としたソフトウェアツールである。τ-Argus は、μ-Argus と同じ 2 つの技術、すなわち、大域的再符号化と局所秘匿を利用する。τ-Argus の局所秘匿とは、統計表のセル秘匿に相当する。
Attribute disclosure: Attribute disclosure is **attribution** independent of identification. This form of disclosure is of primary concern to **NSIs** involved in **tabular data** release and arises from the presence of empty cells either in a released table or linkable set of tables after any subtraction has taken place. Minimally, the presence of a single zero within a table means that an intruder may infer from mere knowledge that a population unit is represented in the table and that the intruder does not possess the combination of attributes within the cell containing the zero.	**属性開示：**属性開示は、識別とは別の問題である。この開示形式は、国家統計機関が集計データ公表において最も懸念するものであり、公表された統計表、若しくは、何らかの減法の行われた後の統計表の組み合わせにおける空白セルの存在から生じる。統計表の中の空白セルの存在は、最小限、侵入者はある母集団ユニットが表中に存在するというだけのわずかな知識から推測でき、その侵入者は空白セルの属性の組み合わせを持たないことを意味する。

Attribution: Attribution is the association or disassociation of a particular attribute with a particular population unit.	**属性：**属性は、特定の母集団ユニットの特定の属性との関連性の有無である。
B	**B**
Barnardisation: A method of disclosure control for tables of counts that involves randomly adding or subtracting 1 from some cells in the table.	**バーナーディゼイション：**統計表において、いくつかのセルにランダムに１を加えるか、引くことで開示抑制を行う手法。
Blurring: Blurring replaces a reported value by an average. There are many possible ways to implement blurring. Groups of records for averaging may be formed by matching on other variables or by sorting on the variable of interest. The number of records in a group (whose data will be averaged) may be fixed or random. The average associated with a particular group may be assigned to all members of a group, or to the "middle" member (as in a moving average). It may be performed on more than one variable with different groupings for each variable.	**ぼかし：**ぼかしとは、報告値を平均値に置き換えることである。ぼかしを実施するための方法は数多くある。平均値を求めるためのレコードのグループを、他の複数の変数に基づくマッチング、あるいは当該変数でソートすることにより形成できる。（平均値を求める）1 つのグループに含まれるレコード数は、固定的あるいはランダムのいずれも可能である。グループの平均値は、グループ内の全メンバー、あるいは（移動平均のような場合には）" 中位 " メンバーに割り当てることができる。各変数について異なるグルーピングにより、複数の変数にぼかしを行なうことができる。
Bottom coding: See **top and bottom coding**.	**ボトム・コーディング：**トップ及びボトム・コーディング参照のこと。
Bounds: The range of possible values of a cell in a table of frequency counts where the cell value has been perturbed or suppressed. Where only margins of tables are released it is possible to infer bounds for the unreleased joint distribution. One method for inferring the bounds across a table is known as the Shuttle algorithm.	**境界：**集計表において撹乱あるいは秘匿されたセルの取りうる値の範囲。表の合計が公表されている場合に限り、未公表の同時分布の境界を推測することが可能である。集計表における境界を推測する一手法として、シャトルアルゴリズムが知られている。
C	**C**
Calculated interval: The interval containing possible values for a suppressed cell in a table, given the table structure and the values published.	**推定区間：**公開された集計表の構造や値が所与の場合、秘匿されたセルの取り得る値の区間。
Cell suppression: In tabular data the cell suppression SDC method consists of **primary and complementary (secondary) suppression**. Primary suppression can be characterised as withholding the values of all risky cells from publication, which means that their value is not shown in the table but replaced by a symbol such as '×' to indicate the suppression. According to the definition of risky cells, in frequency count tables all cells containing small counts and in tables of magnitudes all cells containing small counts or presenting a case of **dominance** have to be primary suppressed. To reach the desired protection for	**セル秘匿：**集計表におけるセル秘匿の手法には、一次及び二次のセル秘匿がある。一次秘匿は、開示リスクのある全セルの値が公表されないようにすることと性格づけすることができる。それは、その値を集計表上で表示せず、秘匿箇所を示す ' × ' 等のシンボルと置き換えることを意味している。開示リスクのあるセルの定義に従って、集計表では、小さい値や占有性を含んでいる全てのセルにおいて一次秘匿が行われなければならない。開示リスクのあるセルを十分に保護するために、補足的（二次的）秘匿と呼ばれる開示リスクのないセルを追加的に秘匿しなけれ

risky cells, it is necessary to suppress additional non-risky cells, which is called **complementary (secondary) suppression**. The pattern of complementary suppressed cells has to be carefully chosen to provide the desired level of ambiguity for the risky cells with the least amount of suppressed information.	ばならない。補足秘匿セルのパターンは、開示リスクのあるセルの秘匿情報量を最小とし、望ましいレベルの曖昧度で保護するために、十分に注意して選択されなければならない。
Complementary suppression: Synonym of **secondary suppression**.	**補足的秘匿：**二次秘匿と同義
Complete disclosure: Synonym of **exact disclosure**.	**完全開示：**exact disclosure を参照のこと。
Concentration rule: Synonym of **(n, k) rule**.	**集中度ルール：**（n,k）ルールと同義
Confidentiality edit: The confidentiality edit is a procedure developed by the U.S. Census Bureau to provide protection in data tables prepared from the 1990 Census. There are two different approaches: one was used for the regular Census data; the other was used for the long-form data which were filled by a sample of the population. Both techniques apply statistical disclosure limitation techniques to the microdata files before they are used to prepare tables. The adjusted files themselves are not released; they are used only to prepare tables. For the regular Census microdata file, the confidentiality edit involves "data swapping" or "switching" of attributes between matched records from different geographical units. For small blocks, the Census Bureau increases the sampling fraction. After the microdata file has been treated in this way, it can be used directly to prepare tables and no further disclosure analysis is needed. For long form data, sampling provides sufficient confidentiality protection, except in small geographic regions. To provide additional protection in small geographic regions, one household is randomly selected and a sample of its data fields are blanked and replaced by imputed values.	**機密保持エディット：**機密保持エディットは、1990 年人口センサスの集計表データを保護するために米国センサス局が開発した手法である。機密保持エディットには異なる 2 種類がある。一つは、10 年に 1 度の定期センサス・データに利用されるものであり、もう一つは、母集団の中から選ばれたサンプルが記入するロング・フォームに利用されるものである。どちらの場合も、表を作成する前に、ミクロデータのファイルに対して開示制限技法を適用する。調整したファイルは、それ自身は公表用にされず、表を作成するためだけに利用される。通常の人口センサスのミクロデータファイルでは機密を保持するために、別の地域のユニットで合致するレコードをデータ・スワッピングあるいはスイッチングに利用している。センサス局では、小ブロックの抽出率を高くしている。ミクロデータ・ファイルに、このような処理を行った後、表の作成に直接利用することで、さらに開示に関する分析を行う必要がなくなる。ロング・フォーム・データでは、小地域区分の場合を除き、サンプリングによって、この保護で十分であることが明らかになった。小地域区分に関しては、追加的保護として、一世帯を無作為に選び、その標本の項目値をブランクにし、補定値と置き換える。
Controlled rounding: To solve the additivity problem, a procedure called controlled rounding was developed. It is a form of **random rounding**, but it is constrained to have the sum of the published entries in each row and column equal to the appropriate published marginal totals. Linear programming methods are used to identify a controlled rounding pattern for a table.	**コントロールされた丸め法：**加法性の問題を解決するために、コントロールされた丸め法という方法が開発された。これは、ランダム丸め法の一種であるが、横列及び縦列の合計が、適正に公表された周辺和と等しくなるよう制約されている。コントロールされた丸め法を行うために、線形計画法が用いられている。
Controlled Tabular Adjustment (CTA): A method to protect tabular data based on the selective adjustment of cell values. Sensitive cell	**コントロールされた集計調整**（CTA）：セル値を選択的に調整することで集計表データを保護する一手法。統計表の加算性を保持す

values are replaced by either of their closest safe values and small adjustments are made to other cells to restore the table additivity. Controlled tabular adjustment has been developed as an alternative to cell suppression.	るために、センシティブなセル値は、最も近くの安全な数値のいずれかに置き換えられ、他のセルには、小さい調整が施される。CTA は、セル秘匿の代替手法として開発されたものである。
Conventional rounding: A disclosure control method for tables of counts. When using conventional rounding, each count is rounded to the nearest multiple of a fixed base. For example, using a base of 5, counts ending in 1 or 2 are rounded down and replaced by counts ending in 0 and counts ending in 3 or 4 are rounded up and replaced by counts ending in 5. Counts ending between 6 and 9 are treated similarly. Counts with a last digit of 0 or 5 are kept unchanged. When rounding to base 10, a count ending in 5 may always be rounded up, or it may be rounded up or down based on a rounding convention.	**伝統的丸め法：**集計表のための開示抑制法の一つ。 伝統的丸め法を利用した場合、各計数は、最も近い定数の倍数に丸められる。例えば、５を基準にした場合、１あるいは２で終わる計数は切り捨て、0 で終わる計数に置き換えられ、３あるいは４で終わる計数は、切り上げ、５で終わる計数に置き換えられる。末尾が６～９でも同様に扱われる。最後の桁が、5 あるいは０で終わる計数は、変更されない。10 を基準に丸める場合、５で終わる計数は、必ず切り上げか、あるいは、丸めの慣行に従い、切り上げか切り捨てられる。
D	**D**
Data divergence: The sum of all differences between two datasets (data-data divergence) or between a single dataset and reality (data-world divergence). Sources of data divergence include: data ageing, response errors, coding or data entry errors, differences in coding and the effect of disclosure control.	**データの相違：**二つのデータセット（データとデータ）間の相違、あるいは、一つのデータセットと現実（データと世界）間の相違の総量。データの相違となるものには、データの老朽化、回答誤差、格付け誤りあるいはデータ入力エラー、格付けにおける相違及び開示抑制効果が含まれている。
Data intruder: A data user who attempts to disclose information about a population unit through **identification** or **attribution**.	**データ侵入者：**識別子あるいは属性を通して母集団ユニットの情報開示を試みようとするデータ・ユーザー。
Data intrusion detection: The detection of a **data intruder** through their behaviour. This is most likely to occur through analysis of a pattern of requests submitted to a **remote data laboratory**. At present this is only a theoretical possibility, but it is likely to become more relevant as **virtual safe settings** become more prevalent.	**データ侵入検出：**データ侵入者をその行為により探知すること。データ侵入は、データ・リモート・ラボラトリーに提出された要求パターンを分析することで、最も探知されるようである。現在のところ、これは理論上の可能性のみであるが、仮想的に安全な環境の見込みが立てば、より一般的になるに従い、より現実味な意味を持つものになるだろう。
Data Intrusion Simulation (DIS): A method of estimating the probability that a **data intruder** who has matched an arbitrary unit against a sample unique in a target microdata file has done so correctly.	**データ侵入シュミレーション（DIS）：**データ侵入者が、任意の母集団ユニットと攻撃対象のミクロデータファイルの標本一意を正しく合致させる確率を推計する一手法。
Data protection: Data protection refers to the set of privacy-motivated laws, policies and procedures that aim to minimise intrusion into respondents' privacy caused by the collection, storage and dissemination of personal data.	**データ保護：**データ保護は、個人データの収集、保管及び公表を原因とする回答者のプライバシーへの侵入の被害を最小にとどめることを目的とした一連のプライバシー保護法、政策及び手続きを表す。

Data swapping: A disclosure control method for microdata that involves swapping the values of variables for records that match on a representative key. In the literature this technique is also sometimes referred to as "multidimensional transformation". It is a transformation technique that guarantees (under certain conditions) the maintenance of a set of statistics, such as means, variances and univariate distributions.	**データ・スワッピング：**代表となるキーが一致したレコードの変数値を交換するミクロデータ開示抑制法。文献では、この技術は、「多次元変換」とも呼ばれている。これは、平均、分散、一変量分布のような一連の統計量を（ある条件で）保証する変換技法である。
Data utility: A summary term describing the value of a given data release as an analytical resource. This comprises the data's **analytical completeness** and its **analytical validity**. Disclosure control methods usually have an adverse effect on data utility. Ideally, the goal of any disclosure control regime should be to maximise data utility whilst minimizing disclosure risk. In practice disclosure control decisions are a trade-off between utility and **disclosure risk**.	**データ有効性：**所与の公表データの分析用リソースとしての価値を記述する要約的用語。これは、データ分析の完全性やその分析の妥当性を含んでいる。開示抑制法は、通常、データ有効性に悪影響がある。どの開示抑制形態の目的も、理想的には、開示リスクを最小にする一方で、データ有効性を最大にすることにある。実際には、開示抑制は、有効性と開示リスク間のトレードオフで決定される。
Deterministic rounding: Synonym of **conventional rounding**.	**決定論的丸め法：**伝統的丸め法と同義
Direct identification: Identification of a statistical unit from its **formal identifiers.**	**直接識別：**正識別子から客体を識別すること。
Disclosive cells: Synonym of **risky cells**.	**開示セル：**開示リスクのあるセルと同義
Disclosure: Disclosure relates to the inappropriate attribution of information to a data subject, whether an individual or an organization. Disclosure has two components: **identification** and **attribution**.	**開示：**開示は、客体である個人、若しくは、機関とその情報（の管理）が不適切であったために結び付けられてしまうことである。開示には、特定（識別）と属性の２つの要素がある。
Disclosure by fishing: This is an attack method where an intruder identifies risky records within a target data set and then attempts to find population units corresponding to those records. It is the type of disclosure that can be assessed through a **special uniques analysis**.	**フィッシングによる開示:**これは、侵入者が、攻撃対象のデータセットから開示リスクのあるレコードを識別し、そのレコードと対応する母集団ユニットを見出そうとする攻撃手法である。特別一意分析から評価できる開示形態である。
Disclosure by matching: Disclosure by the linking of records within an identification dataset with those in an anonymised dataset.	**マッチングによる開示：** 識別データセットのレコードと匿名化データセット・レコードを結びつけることによる開示。
Disclosure by response knowledge: This is disclosure resulting from the knowledge that a person was participating in for a particular survey. If an intruder knows that a specific individual has participated in the survey, and that consequently his or her data are in the data set, identification and disclosure can be accomplished more easily.	**回答に関する知識による開示：**これは、ある人物が特定の調査の対象となったことを知っていた結果起こる開示である。特定の個人が調査に参加し、その結果、データセット内に含まれることを、侵入者が知っていれば、当該個人の識別及び開示はより容易となる。
Disclosure by spontaneous recognition: This	**偶発的認知による開示：**これは、データセッ

means the recognition of an individual within the dataset. This may occur by accident or because a data intruder is searching for particular individual. This is more likely to be successful if the infividual has a rare combination of characteristics which is known to the intruder.	ト内の個人の認知を意味する。これは、偶発的、あるいは、データ侵入者がある特定の人物を調べていることから起こる可能性がある。その個人を特定できるような特徴が、希少であることを侵入者が知っている場合、開示される可能性が高くなる。
Disclosure control methods: There are two main approaches to control the disclosure of confidential data. The first is to reduce the information content of the data provided to the external user. For the release of tabular data this type of technique is called **restriction baesd disclosure control method** and for the release of microdata the expression disclosure control by data reduction is used. The second is to change the data before the dissemination in such a way that the disclosure risk for the confidential data is decreased, but the information content is retained as much as possible. These are called **perturbation based disclosure control methods**.	**開示抑制法：**機密データの開示を抑制するためには主に 2 つのアプローチがある。まず、外部ユーザーに提供されるデータの情報コンテンツを減らす方法である。このような方法は、集計表データを公表する場合には、制限的開示抑制法と呼ばれ、ミクロデータの場合には、データ削減による開示抑制という表現が使用される。第 2 の方法は、機密データの開示リスクを減らしつつ、できるだけ多くの情報を含有するような方法で、データ公開前にデータを改変する方法である。これらは撹乱的開示抑制法と呼ばれる。
Disclosure from analytical outputs: The use of output to make attributions about individual population units. This situation might arise to users that can interrogate data but do not have direct access to them such as in a **remote data laboratory**. One particular concern is the publication of residuals.	**アウトプットの分析による開示：**個別の母集団ユニットを特定するためのアウトプットの利用。この状況は、リモート・データ・ラボラトリー内のように、データに指示することはできるが、直接アクセスしないユーザーに起こる可能性がある。特殊な問題として、残差の公開がある。
Disclosure limitation methods: Synonym of disclosure control methods.	**開示制限法：**開示抑制法と同義
Disclosure risk: A disclosure risk occurs if an unacceptably narrow estimation of a respondent's confidential information is possible or if exact disclosure is possible with a high level of confidence.	**開示リスク：**開示リスクは、回答者の機密情報が受け入れがたいほど狭い範囲で推計されるか、あるいは、高度な機密に関して、完全開示が行われる可能性がある場合に起こる。
Disclosure scenarios: Depending on the intention of the intruder, his or her type of a priori knowledge and the microdata avialbable, three different types of disclosure or disclosure scenarios are possible for microdata: **disclosure by matching**, **disclosure by response knowledge** and **disclosure by spontaneous recognition**.	**開示シナリオ：**侵入者の意図、事前の知識、及び入手可能なミクロデータによって、3 つの異なる種類の開示あるいは開示シナリオが考えられる。すなわち、マッチングによる開示、回答に関する知識による開示、偶発的認知による開示である。
Dissemination: Supply of data in any form whatever: publication, access to databases, microfiches, telephone communications, etc.	**公開：**データ提供を意味し、出版、データベースへのアクセス、マイクロフィッシュ、電話による問合せなど、その方法の如何を問わない。
Disturbing the data: This process involves changing the data in some systematic fashion, with	**データの撹乱：**これは、その処理によって個別客体の開示情報の正確性としては不十分

the result that the figures are insufficiently precise to disclose information about individual cases.	となるよう何らかの系統的な方法でデータを変更することである。
Dominance rule: Synonym of **(n,k) rule**.	**占有ルール：**（n,k）ルールと同義
E	**E**
Exact disclosure: Exact disclosure occurs if a user is able to determine the exact attribute for an individual entity from released information.	**完全開示：**完全開示は、ユーザーが公開された情報から個々の客体の正確な属性を特定できる場合に発生する。
F	**F**
Formal identifier: Any variable or set of variables which is structurally unique for every population unit, for example a population registration number. If the formal identifier is known to the intruder, identification of a target individual is directly possible for him or her, without the necessity to have additional knowledge before studying the microdata. Some combinations of variables such as name and address are pragmatic formal identifiers, where non-unique instances are empirically possible,but with negligible probability.	**正識別子：**どの母集団ユニットにおいても構造的に一意である変数あるいは一連の変数、例えば住民登録ナンバーである。もし正識別子が、侵入者に知られている場合には、ターゲットの個人の識別は直接的に可能であり、ミクロデータを調べる前に他の知識を持つ必要がない。経験的に一意ではない例もありうるが、氏名、住所のような変数の組み合わせは、実際のところ無視できない確率で正識別子となる。
G	**G**
Global recoding: Problems of confidentiality can be tackled by changing the structure of data. Thus, rows or columns in tables can be combined into larger class intervals or new groupings of characteristics. This may be a simpler solution than the suppression of individual items, but it tends to reduce the descriptive and analytical value of the table. This protection technique may also be used to protect microdata.	**大域的再符号化（再格付け）：** 機密保持の問題は、データの構成を変更することにより取り組みが可能である。表の縦列あるいは横列を、より大きな分類に統合したり、新たな分類区分を用いる方法である。これは、個々の項目の秘匿よりも簡単な解決方法であるが、表の記述的、分析的価値を減じる傾向がある。この保護技法は、ミクロデータの保護にも利用することが可能である。
H	**H**
HITAS: A heuristic approach to **cell suppression** in hierarchical tables.	**HITAS：**階層的な統計表におけるセル秘匿のための発見的解決法
I	**I**
Identification: Identification is the association of a particular record within a set of data with a particular population unit.	**特定（識別）：** 特定（識別）は、一連のデータの特定レコードを特定の母集団ユニットと関連付けること。
Identification dataset: A dataset that contains formal identifiers.	**識別データセット：**正識別子を含むデータセット
Identification data: Those personal data that allow direct identification of the data subject, and which are needed for the collection, checking and matching of the data, but are not subsequently used	**識別データ：**客体の直接識別を可能にする個人データ。データの収集、チェック、照合のために必要であるが、その後の統計結果を作成するときには使われない。

for drawing up statistical results.	
Identification key: Synonym of **key**.	**識別キー：**キーと同義
Identification risk: This risk is defined as the probability that an intruder identifies at least one respondent in the disseminated microdata. This identification may lead to the disclosure of (sensitive) information about the respondent. The risk of identification depends on the number and nature of quasi-identifiers in the microdata and in the a priori knowledge of the investigator.	**識別リスク：**識別リスクの定義は、侵入者が、公開されたミクロデータの中から少なくとも 1 名の回答者を識別するリスクの確率である。この識別により、回答者の（センシティブな）情報を開示する可能性がある。識別リスクは、ミクロデータ及び侵入者の事前知識における準識別子の数と性質により決まる。
Identifying variable: A variable that either is a formal identifier or forms part of a formal identifier.	**識別変数：**正識別子か、あるいは、正識別子の一部を形作る変数
Indirect identification: Inferring the identity of a population unit within a microdata release other than from **direct identification**.	**間接識別：**ミクロデータの公開において直接識別以外で母集団ユニットの識別を推測すること。
Inferential disclosure: Inferential disclosure occurs when information can be inferred with high confidence from statistical properties of the released data. For example, the data may show a high correlation between income and purchase price of home. As the purchase price of a home is typically public information, a third party might use this information to infer the income of a data subject. In general, NSIs are not concerned with inferential disclosure for two reasons. First, a major purpose of statistical data is to enable users to infer and understand relationships between variables. If NSIs equated disclosure with inference, no data could be released. Second, inferences are designed to predict aggregate behaviour, not individual attributes, and thus often poor predictors of individual data values.	**推論開示：**公開されたデータの統計的性質から高い確度で情報が推論できるときに、推論開示が発生する。例えば、データが所得と住宅購入価格に強い相関関係があることを示したとする。住宅購入価格は典型的な公開情報であるため、第三者は、この情報を使って、客体の収入を推論することができる。一般的に、国家統計機関は、次の 2 つの理由から、推論開示に関心を持たない。第 1 に、統計データの主要目的は、ユーザーが、変数間の関係を推論し理解することを可能にすることである。もし国家統計機関が、開示と推論を同等とみなしたとすると、いかなるデータも公開することができなくなる。第 2 に、推論は、集合的行動を予測することを意図したものであり、個人の属性を明らかにするものでない。従って、個人データの予測手段としては不十分であることが多い。
Informed consent: Basic ethical tenet of scientific research on human populations. Sociologists do not involve a human being as a subject in research without the informed consent of the subject or the subject's legally authorized representative, except as otherwise specified. Informed consent refers to a person's agreement to allow personal data to be provided for research and statistical purposes. Agreement is based on full exposure of the facts the person needs to make the decision intelligently, including awareness of any risks involved, of uses and users of the data, and of alternatives to providing the data.	**インフォームド・コンセント：**人間に関する科学的調査研究の基本倫理原則。社会学では、法律に別段の定めがある場合を除いて、調査対象者あるいは調査対象者の法律上の代表者のインフォームド・コンセントなしに、その人間を、調査研究の対象としない。インフォームド・コンセントは、統計調査の目的で個人データの提供を許可することに同意することをいう。同意は、個人が意思決定をしたという明確な事実に基づかなければならない。その中には、内在するあらゆるリスク、データの利用やデータの利用者、データ提供に対する代替手段を知ることも含まれている。

Intruder: A data user who attempts to link a respondent to a microdata record or make attributions about particular population units from aggregate data. Intruders may be motivated by a wish to discredit or otherwise harm the NSI, the survey or the government in general, to gain notoriety or publicity, or to gain profitable knowledge about particular respondents.	**侵入者：** 回答者とミクロデータ・レコードを結び付けたり、合計値から母集団ユニットを特定しようと試みるデータ・ユーザー。侵入者の動機は、NSI、調査、あるいは政府一般の信用をおとしめたり、傷つけたり、悪い評判を立てたり、公共、若しくは、特定の回答者に関する有益な情報を得ようと侵入を図るかもしれない
J	**J**
K	**K**
Key: A set of **key variables**.	**キー：**一連のキー変数
Key variable: A variable in common between two datasets, which may therefore be used for linking records between them. A key variable can either be a **formal identifier** or a **quasi-identifier**.	**キー変数：**２つのデータセットに共通する変数で、データセット間のレコードを結び付ける目的で利用される可能性がある。キー変数は、正識別子、あるいは、準識別子のどちらかである。
L	**L**
Licensing agreement: A permit, issued under certain conditions, for researchers to use confidential data for specific purposes and for specific periods of time. This agreement consists of contractual and ethical obligations, as well as penalties for improper disclosure or use of identifiable information. These penalties can vary from withdrawal of the license and denial of access to additional data sets to the forfeiting of a deposit paid prior to the release of a **microdata** file. A licensing agreement is almost always combined with the signing of a contract. This contract includes a number of requirements: specification of the intended use of the data; instruction not to release the **microdata** file to another recipient; prior review and approval by the releasing agency for all user outputs to be published or disseminated; terms and location of access and enforceable penalties.	**ライセンス協定：**所定の条件の下で、研究者に対し、特定の目的及び特定の期間において機密データの利用を許可すること。この協定には、契約上の義務と倫理義務、識別可能な情報の不適切な開示及び利用に対する罰則から成る。ライセンス協定に定める規則に違反した場合の罰則は、ライセンスの停止、データセットへのアクセス禁止から、ミクロデータファイル開示前の預託金の没収まで様々である。ライセンス協定は、ほとんどの場合、契約の署名が必要である。この契約には、幾つかの要件が盛り込まれる。すなわち、データの使用目的の明示、ミクロデータファイルを他人に公開しないという指示、研究成果の公開・配布前に所管機関の事前審査と承認を受けること、データへのアクセスの条件と場所、執行可能な罰則などである。
Local recoding: A disclosure control technique for microdata where two (or more) different versions of a variable are used dependent on some other variable. The different versions will have different levels of coding. This will depend on the distribution of the first variable conditional on the second. A typical example occurs where the distribution of a variable is heavily skewed in some geographical areas. In the areas where the distribution is skewed minor categories may be combined to produce coarser variable.	**局所的再符号化（再格付け）：**いくつかの他の変数に依存した２つ（あるいはそれを超える）の区分の異なる変数を使ったミクロデータ開示抑制技術。区分が違うと、格付けのレベルも異なり、第２の変数に対する第１の変数の分布の条件に依存している。典型的な例として、変数の分布が、地域によって大きく歪んでいる場合に生じる。分布が歪んでいる地域では、小さな区分は組み合わされ、粗い変数を作るかもしれない。

Local suppression: Protection technique that diminishes the risk of recognition of information about individuals or enterprises by suppressing individual scores on identifying variables.	**局所秘匿：**識別変数について個別の値を秘匿することにより、個人あるいは企業の情報が認識されるリスクを減じる保護技法。
M	**M**
Macrodata: Synonym of **tabular data**.	**マクロデータ：**集計表データと同義
Micro aggregation: Records are grouped based on a proximity measure of variables of interest, and the same small groups of records are used in calculating aggregates for those variables. The aggregates are released instead of the individual record values.	**ミクロ（マイクロ）・アグリゲーション：**レコードは、関係するすべての変数の類似度に基づきグループ分けされ、同一小グループごとに、各変数の合計値が計算される。公開されるのは個々のレコード値ではなく合計値である。
Microdata: A microdata set consists of a set of records containing information on individual respondents or on economic entities.	**ミクロ（マイクロ）データ：**ミクロデータ・セットは、個々の回答者、若しくは、経済主体についての情報を含んだレコードのセットにより構成される。
Minimal unique: A combination of variable values that are unique in the **microdata** set at hand and contain no proper subset with this property (so it is a minimal set with the uniqueness property).	**最小一意：**当該のミクロデータセットの中で一意で、この特性を持つ適当なサブセットを含まない変数値の組み合わせ（そのため一意の特性としては最小セット）。
N	**N**
NSI(s): Abbreviation for National Statistical Institute(s).	**NSI(s)**：国家統計機関（National Statistical Institute(s)）の略語
(n,k) rule: A cell is regarded as confidential, if the n largest units contribute more than k % to the cell total, e.g. n=2 and k=85 means that a cell is defined as risky if the two largest units contribute more than 85 % to the cell total. The n and k are given by the statistical authority. In some **NSIs** the values of n and k are confidential.	**（n,k）ルール：**（同一セルに含まれる客体のうち）最も値の大きいn個の客体が、セルの値のk%を上回っている場合に、そのセルは機密セルとみなすルール。例えば、nが2、kが85の場合には、2個の最も値の大きい客体が、セルの値の85%を上回っている場合に、そのセルは機密セルとみなされる。nとkは、統計機関が決定しているが、統計機関によってはn及びkの値を機密にしている。
O	**O**
On-site facility: Facility that has been established on the premises of several NSIs. It is a place where external researchers can be permitted access to potentially disclosive data under contractual agreements which cover the maintenance of confidentiality, and which place strict controls on the uses to which the data can be put. The on-site	**オンサイト施設：**複数の国家統計機関が設置している施設。これは、外部の研究者が、契約条件に基づき潜在的な開示データにアクセスすることができる場所で、機密が保持され、データの利用は厳格に管理されている。オンサイト施設は、機密データの分析が可能な“安全な環境”と捉えることができる。オ

facility can be seen as a 'safe setting' in which confidential data can be analysed. The on-site facility itself would consist of a secure hermetic working and data storage environment in which the confidentiality of the data for research can be ensured. Both the physical and the IT aspects of security would be considered here. The on-site facility also includes administrative and support facilities to external users, and ensures that the agreed conditions for access to the data were complied with.	ンサイト施設自体が、研究用データの機密を保証できる安全で密閉された作業環境及びデータ保存環境を有している。ここでは、物理的及び IT の側面のセキュリティが考慮されている。オンサイト施設には、外部ユーザーのための管理及びサポート設備も備わり、データアクセスに関する合意条件が履行されたことを保証する。
Ordinarily rounding: Synonym of **conventional rounding**.	**通常の丸め法：**伝統的丸め法と同義
Oversuppression: A situation that may occur during the application of the techniques of cell suppression. This denotes the fact that more information has been suppressed than strictly necessary to maintain confidentiality.	**過剰秘匿：**セルの秘匿を行っているときに発生する可能性のある状況。これは、機密保持のために必要以上に情報の秘匿が行われたことを示す言葉である。
P	**P**
Partial disclosure: Synonym of **approximate disclosure**.	**部分開示：**近似的開示と同義
Passive confidentiality: For foreign trade statistics, EU countries generally apply the principle of "passive confidentiality", that is they take appropriate measures only at the request of importers or exporters who feel that their interests would be harmed by the dissemination of data.	**受動的機密保持：**外国貿易統計の場合、EU 諸国は一般的に" 受動的機密保持 "の原則を採用している。すなわち、データ公開により自らの利益が損なわれると感じる輸入業者あるいは輸出業者からの要請があった場合のみ、適切な対策を講じるというものである。
Personal data: Any information relating to an identified or identifiable natural person ('data subject'). An identifiable person is one who can be identified, directly or indirectly. Where an individual is not identifiable, data are said to be anonymous.	**個人データ：**識別された、若しくは、識別可能な客体に関する情報。識別可能な者とは、直接的、若しくは、間接的に、識別可能な者を言う。個人の識別が不可能な場合には、そのデータは匿名データと言われる。
Perturbation based disclosure control methods: Techniques for the release of data that change the data before the dissemination in such a way that the disclosure risk for the confidential data is decreased but the information content is retained as far as possible. Perturbation based methods falsify the data before publication by introducing an element of error purposely for confidentiality reasons. For example, an error can be inserted in the cell values after a table is created, which means that the error is introduced to the output of the data and will therefore be referred to as output perturbation. The error can also be	**撹乱的開示抑制法：**データの公開前に、機密データの開示リスクを減じる一方で、情報のコンテンツをできるだけ維持するような方法で、データを改変するデータの公表技法。撹乱的な方法とは、機密保持のために意図的にエラーの要素を導入してデータを改変する方法である。例えば、エラーは表が作成された後に、セルの数値に挿入することが考えられる。すなわち、データのアウトプットにエラーを挿入する。この場合アウトプット撹乱と言う。エラーは、ミクロデータ・レベルのオリジナルデータ、すなわち表作成の前にインプット・データに挿入することもでき

inserted in the original data on the **microdata** level, which is the input of the tables one wants to create; the method will then be referred to as data perturbation – input perturbation being the better but uncommonly used expression. Possible perturbation methods are: - Rounding; - perturbation, for example, by the addition of random noise or by the **post randomization Method**; - disclosure control methods for micro-data applied to tabular data.	る。この方法を、データ撹乱と言う。インプット撹乱の方が好ましい表現であるが、一般的でない。可能な撹乱方法は次のとおりである。 ・ 丸め法。 ・ 撹乱、例えば、ランダムにノイズを加えたり、または事後ランダム化法（PRAM）によって。 ・ 集計データに応用したミクロデータの開示抑制法。
. Population unique: A record within a dataset which is unique within the population on a given key.	**母集団一意：**所与のキーに関して母集団内で一意である（データセットの中の）レコード
P-percent rule: A (p,q) rule where q is 100 %, meaning that from general knowledge any respondent can estimate the contribution of another respondent to within 100 % (i.e., knows the value to be nonnegative and less than a certain value which can be up to twice the actual value).	**P パーセント・ルール：**q が 100 パーセントである(p,q)ルール。ｑが、100 パーセントであるということは、一般的な知識からある回答者が、別の回答者の寄与率を 100 パーセント以内で推計することができることを意味する（すなわち、値が非負で、実際の値の２倍を限度に、それ以下であることが知られている）。
(p,q) rule: It is assumed that out of publicly available information the contribution of one individual to the cell total can be estimated to within q per cent (q=error before publication); after the publication of the statistic the value can be estimated to within p percent (p=error after publication). In the (p,q) rule the ratio p/q represents the information gain through publication. If the information gain is unacceptable the cell is declared as confidential. The parameter values p and q are determined by the statistical authority and thus define the acceptable level of information gain. In some **NSIs** the values of p and q are confidential.	**（p,q）ルール：**一般に入手可能な情報から、セルの値に対して個々の客体の寄与率がｑパーセント（ｑ＝公表前の誤差）以内で推定できると想定した場合に、統計発表後の推定精度がｐパーセント（ｐ＝公表後の誤差）以内に入るか否かで機密セルを判定するルール。（p,q）ルールの場合、p,q の比率が、公表により獲得される情報となる。この獲得される情報が許容不能な場合に、そのセルは機密セルとなる。統計機関がパラメータ値 p と q、したがって獲得情報の許容レベルを決定する。統計機関によっては、p 及び q の値を、機密にしている。
Post Randomisation Method (PRAM): Protection method for microdata in which the scores of a categorial variable are changed with certain probabilities into other scores. It is thus intentional misclassification with known misclassification probabilities.	**事後ランダム化法（PRAM）：**カテゴリー変数のスコアに、ある確率を付加しスコアを変化させることでミクロデータを保護する手法。既知の誤分類確率を持つ意図的誤分類である。
Primary confidentiality: It concerns tabular cell data, whose dissemination would permit attribute disclosure. The two main reasons for declaring data to be primary confidential are: - too few units in a cell; - dominance of one or two units in a cell. The limits of what constitutes “too few” or “dominance” vary between statistical domains.	**一次機密：**これは、その公開により属性開示が可能となる集計表のセルデータを言う。一次機密と判定する場合は、主に２つある。 ・ セル内の客体数（件数）が少な過ぎる。 ・ 一つあるいは二つの客体の占有度が高い。 “少な過ぎる”あるいは“占有度が高い”限度は、統計分野により異なる。

Primary protection: Protection using disclosure control methods for all cells containing small counts or cases of dominance.	**一次保護：**開示抑制法を用いて、件数が少ないあるいは占有度が高い全てのセルを、開示保護すること。
Primary suppression: This techniques can be characterized as withholding all disclosive cells from publication, which means that their value is not shown in the table, but replaced by a symbol such as "x" to indicate the suppression. According to the definition of disclosive cells, in frequency count tables all cells containing small counts and in tables of magnitudes all cells containing small counts or representing cases of dominance have to be primary suppressed.	**一次秘匿：**開示セルを全て公表から除外すること、すなわち、これらのセルの数値を表中に掲げず、秘匿したことを示す " x " などのマークで置き換える技法。開示セルの定義に従い、度数表においては件数の少ないすべてのセル、数量表においては件数の少ないセルあるいは（少数客体の）占有度の高いセルはすべて、一次秘匿しなければならない。
Prior-posterior rule: Synonym of the **(p,q) rule**.	**事前事後ルール：**（p,q）ルールと同義
Privacy: Privacy is a concept that applies to data subjects while confidentiality applies to data. The concept is defined as follows: "It is the status accorded to data which has been agreed upon between the person or organisation furnishing the data and the organization receiving it and which describes the degree of protection which will be provided." There is a definite relationship between confidentiality and privacy. Breach of confidentiality can result in disclosure of data which harms the individual. This is an attack on privacy because it is an intrusion into a person's self-determination on the way his or her personal data are used. Informational privacy encompasses an individual's freedom from excessive intrusion in the quest for information and an individual's ability to choose the extent and circumstances under which his or her beliefs, behaviours, opinions and attitudes will be shared with or withheld from others.	**プライバシー：**機密がデータに適用されるのに対し、プライバシーは客体に適用される概念である。この概念は次のように定義される。" データを提供する人物あるいは機関とデータを受け取る機関の間で合意されたデータに関するステータスで、実施される保護レベルを示すもの "。機密とプライバシーの間には一定の関係がある。機密保持が履行されないと、データが開示され、個人を害する可能性がある。これは、個人データの利用方法は自身が決定できるという原則を脅かすことになるため、プライバシーの侵害となる。情報プライバシーには、情報を探知しようとする過度な侵害からの個人の自由及び、信念、行動、意見、態度を周囲と共有したり、周囲と切り離す範囲や状況を選択する個人の能力が含まれている。
Probability based disclosure (approximate or exact): Sometimes although a fact is not disclosed with certainty, the published data can be used to make a statement that has a high probability of being correct.	**確率的開示（近似的あるいは完全）：**正確な事実が開示されていない場合でも、公表データから高い確率で正しい情報を推定できることがある。
Q	**Q**
Quasi-identifier: Variable values or combinations of variable values within a dataset that are not structural uniques but might be empirically unique and therefore in principle uniquely identify a population unit.	**準識別子：**構造的に一意ではないが、経験的に一意であるために、原則的に一意に母集団ユニットを識別する（データセットの中の）変数値あるいは変数値の組み合わせ。

R	R
Randomized response: Randomized response is a technique used to collect sensitive information from individuals in such a way that survey interviewers and those who process the data do not know which of two alternative questions the respondent has answered.	**無作為化回答：**回答の無作為化は、2 つの質問のうち、回答者がどちらの質問に回答したかが調査員及びデータ処理者には分からないような方法で、センシティブな情報を個人から収集する技法を言う。
Random perturbation: This is a disclosure control method according to which a noise, in the form of a random value is added to the true value or, in the case of categorical variables, where another value is randomly substituted for the true value.	**ランダム撹乱：**これは、開示抑制法の一つで、表の実際の値に、ランダムな値のノイズを加えるか、あるいはカテゴリー変数の場合、別の数値をランダムに実際の値と置き換える方法である。
Random rounding: In order to reduce the amount of data loss that occurs with suppression, alternative methods have been investigated to protect sensitive cells in tables of frequencies. Perturbation methods such as random and controlled rounding are examples of such alternatives. In random rounding cell values are rounded, but instead of using standard rounding conventions a random decision is made as to whether they will be rounded up or down. The rounding mechanism can be set up to produce unbiased rounded results.	**ランダム丸め法：**秘匿によるデータのロスを抑制するために、度数表の機密セルを保護する代替法が研究されてきた。ランダム丸め法やコントロールされた丸め法などの撹乱法は、その例である。ランダム丸め法の場合は、セルの値はラウンドされるが、標準的な丸め法と異なり、切り上げか切り捨てかをランダムに決定する。この処理法は、表の加算性を損なう。丸めた結果を偏りのないものにするために、丸め法のメカニズムを、調整することができる。
Rank swapping: Rank swapping provides a way of using continuous variables to define pairs of records for swapping. Instead of insisting that variables match (agree exactly), they are defined to be close based on their proximity to each other on a list sorted on the continuous variable. Records which are close in rank on the sorted variable are designated as pairs for swapping. Frequently in rank swapping the variable used in the sort is the one that will be swapped.	**ランク・スワッピング：**ランク・スワッピングは、スワッピングするレコードのペアを連続変数を利用して選択する方法である。変数のマッチング（完全一致）に固執せず、連続性のある変数でソートされたリスト上で相互の近さに基づいて行う。ソート変数によるランクの近い２つのレコードを、スワッピングするペアとして指定する。ランク・スワッピングにおいては、ソートに用いた変数が、値を交換する対象となることが多い。
Record linkage process: Process attempting to classify pairs of matches in a product space AxB from two files A and B into M, the set of true links, and U, the set of non-true links.	**レコード・リンケージ処理：**A と B の 2 つのファイルの積空間 A×B から、真のリンク M と真でないリンク U のペアとに選り分けることを試みる処理。
Record swapping: A special case of **data swapping**, where the geographical codes of records are swapped.	**レコード・スワッピング：**レコードの地理コードが置き換えられたデータ・スワッピングの特別なケース。
Remote access: On-line access to protected microdata.	**リモート・アクセス：**保護が施されたミクロデータへのオンライン・アクセス
Remote data laboratory: A virtual environment providing remote execution facilities.	**リモート・データ・ラボラトリー：**仮想環境下でリモート運用できる施設
Remote execution: Submitting scripts on-line	**リモート実行：**機関によって保護されたネッ

for execution on disclosive microdata stored within an institute's protected network. If the results are regarded as **safe data**, they are sent to the submitter of the script. Otherwise, the submitter is informed that the request cannot be acquiesced. Remote execution may either work through submitting scripts for a particular statistical package such as SAS, SPSS or STATA which runs on the remote server or via a tailor made client system which sits on the user's desktop.	トワーク内に格納された開示ミクロデータを使用するためのオンライン・スクリプトを提出すること。その結果、安全なデータとして許可されれば、データ提出者に返送される。そうでない場合は、提出者にその旨通知される。リモート・サーバー上か、ユーザーのデスクトップ上に設定されたオーダーメイドのクライアントシステムかのどちらかで動く SAS、SPSS あるいは STATA のような特定の統計パッケージのためのスクリプトを提出すればリモートでのデータ使用を行うこともできる。
Residual disclosure: Disclosure that occurs by combining released information with previously released or publicly available information. For example, tables for nonoverlapping areas can be subtracted from a larger region, leaving confidential residual information for small areas.	**残差開示：**以前に公開された情報や、あるいは既に一般的に利用可能になっている情報を組み合わせることで生じる開示。例えば、小地域の情報を秘匿しても、より広い地域から地域的に重なっていない統計表の情報を引くことで、機密にした小地域情報を残りの情報として開示してしまう。
Restricted access: Imposing conditions on access to the **microdata**. Users can either have access to the whole range of raw protected data and process individually the information they are interested in – which is the ideal situation for them – or their access to the protected data is restricted and they can only have a certain number of outputs(e.g. tables) or maybe only outputs of a certain structure. Restricted access is sometimes necessary to ensure that linkage between tables cannot happen.	**制限されたアクセス：**ミクロデータへのアクセスに関して条件を課すこと。ユーザーは、保護されているデータ全体にアクセスし、ユーザーが関心のある情報を個々に処理するか－ユーザーにとっては、これが理想的な状況であるが－あるいは保護されているデータへのアクセスが制限され、所定の数のアウトプット（例えば表）、若しくは、所定の構造のアウトプットのみを入手できる。アクセスの制限は、表間のリンケージで開示が起きることを防ぐために必要となることがある。
Restricted data: Synonym of **safe data**.	**制限されたデータ：**安全なデータ（safe data）と同義
Restriction based disclosure control method: Method for the release of tabular data, which consists in reducing access to the data provided to the external user. This method reduces the content of information provided to the user of the **tabular data**. This is implemented by not publishing all the figures derived from the collected data or by not publishing the information in as detailed a form as would be possible.	**制限的開示抑制法：**集計表データの公表方法で、外部ユーザーに提供するデータへのアクセスを抑制するもの。この手法は、集計表のユーザーに提供される情報コンテンツを抑制するものである。これは、収集データから生じた数字を必ずしもすべて公表しない、あるいは可能な限り詳細な様式による情報を公表しないことで行われる。
Risky cells: The cells of a table which are nonpublishable due to the risk of statistical disclosure are referred to as risky cells. By definition there are three types of risky cells: small counts, dominance and complementary suppression cells.	**開示リスクのあるセル：**統計開示の危険性があるために非公表となる表のセルは、開示リスクのあるセルと呼ばれる。定義によって、小さい値、占有度及び補足的秘匿の三つのタイプの開示リスクのあるセルがある。
Risky data: Data are considered to be disclosive when they allow statistical units to be identified, either directly or indirectly, thereby disclosing	**開示リスクのあるデータ：**客体を直接的か、あるいは間接的に識別できるような個別情報が開示されるとき、データは、開示された

individual information. To determine whether a statistical unit is identifiable, account shall be taken of all the means that might reasonably be used by a third party to identify the said statistical unit.	と考えられる。いわゆる「客体」を識別する第三者は、客体が識別できるかどうかを決めるために、考えられるすべての手法をとるものと考えるべきである。
Rounding: Rounding belongs to the group of disclosure control methods based on output-perturbation. It is used to protect small counts in **tabular data** against disclosure. The basic idea behind this disclosure control method is to round each count up or down either deterministically or probabilistically to the nearest integer multiple of a rounding base. The additive nature of the table is generally destroyed by this process. Rounding can also serve as a **recoding** method for microdata.	**丸め法：**丸め法は、アウトプット撹乱に基づく開示抑制法に属する。この方法は、集計表データの件数の少ないセルを開示から防ぐために利用される。この開示抑制法の背景にある基本的なアイデアは、件数を、丸め基準の整数の倍数のうち、最も近いものに確定的あるいは確率的に切り上げ、若しくは、切り捨てることである。表の加法性の特徴は、一般にこのプロセスで損なわれる。丸め法は、また、ミクロデータの再符号化（再格付け）手法として捉えられる。
R-U map: A graphical representation of the trade off between disclosure risk and data utility.	**R-U マップ：**開示リスクとデータ有効性のトレードオフを図式的に表したもの。
S	**S**
Safe data: **Microdata** or **macrodata** that have been protected by suitable **Statistical Disclosure Control** methods.	**安全なデータ：**適切な統計開示抑制法によって保護されたミクロデータ、あるいは、マクロデータ。
Safe setting: An environment such as a **microdata** lab whereby access to a disclosive dataset can be controlled.	**安全な環境：**開示データセットへのアクセスが管理されているミクロデータ・ラボのような環境。
Safety interval: The minimal **calculated interval** that is required for the value of a cell that does not satisfy the primary suppression rule.	**安全な区間：**一次秘匿ルールには達しないが、計算上でセル値に要求される最小区間のこと。
Sample unique: A record within a dataset which is unique within that dataset on a given **key**.	**標本一意：**所与のキーに関してデータセット内で一意である（データセットの中の）レコード。
Sampling: In the context of disclosure control, this refers to releasing only a small proportion of the original data records on a **microdata** file.	**サンプリング：**開示抑制において、オリジナル・データ・レコードの少数部分のみをミクロデータ・ファイル・レコードとして公表すること。
Sampling fraction: The proportion of the population contained within a data release. With simple random sampling, the sample fraction represents the proportion of population units that are selected in the sample. With more complex sampling methods, this is usually the ratio of the number of units in the sample to the number of units in the population from which the sample is selected.	**標本抽出率：**公表データに含まれる母集団の割合。単純無作為抽出では、標本抽出率は、標本として選択された母集団ユニットの割合を表している。さらに複雑な標本抽出法では、通常、標本を抽出した母集団ユニットに対する標本ユニット数の割合である。
Scenario analysis: A set of pseudo-criminological methods for analysing and	**シナリオの分析：**データ侵入の可能性がある危険なルートを分析、分類する擬似犯罪手

classifying the plausible risk channels for a data intrusion. The methods are based around first delineating the means, motives and opportunity that an intruder may have for conducting the attack. The output of such an analysis is a specification of a set of **keys** likely to be held by **data intruders**.	法。その手法は、まず、侵入者が攻撃を実行する方法、動機及び機会を、詳細に記述することを前提にしている。データ侵入者が得ようとする一連のキーの特定が分析のアウトプットになる。
Secondary data intrusion: After an attempt to match between identification and target datasets an intruder may discriminate between non-unique matches by further direct investigations using additional variables.	**二次的データ侵入：**侵入者は、ターゲットとなるデータセットの識別を試みた後、追加的に変数を用い、さらに深く分析することで、一意でない組み合わせの識別を行う。
Secondary disclosure risk: It concerns data which is not primary disclosive, but whose dissemination, when combined with other data permits the identification of a microdata unit or the disclosure of a unit's attribute.	**二次的開示リスク：**これは、一次的な開示データではないが、その公表によって、他のデータと組み合わせた時にミクロデータの客体や、あるいは客体の属性の識別が可能となるデータを言う。
Secondary suppression: To reach the desired protection for risky cells, it is necessary to suppress additional non-risky cells, which is called secondary suppression or complementary suppression. The pattern of complementary suppressed cells has to be carefully chosen to provide the desired level of ambiguity for the disclosive cells at the highest level of information contained in the released statistics.	**二次秘匿：**開示リスクのあるセルに対して望ましい保護を実現するために、開示リスクのないセルを追加的に秘匿する必要がある。これを二次秘匿あるいは補足的秘匿と言う。補足的な秘匿セルの選択は、公表する統計の中に含まれる情報のレベルを最大化しつつ、開示セルの曖昧度が望ましいレベルになるよう、慎重に行なわなければならない。
Security: An efficient disclosure control method provides protection against exact disclosure or unwanted narrow estimation of the attributes of an individual entity, in other words, a useful techniques prevents exact or partial disclosure. The security level is accordingly high. In the case of disclosure control methods for the release of **microdata** this protection is ensured if the identification of a respondent is not possible, because the identification is the prerequisite for disclosure.	**セキュリティ：**効率的な開示抑制法は、個人の属性の完全開示あるいは好ましくないほど高い精度の推定を防止できる。言い換えれば、有効な技法を講じることにより、完全開示あるいは部分開示を防止することができる。これによりセキュリティレベルは高まる。ミクロデータ公開のための開示抑制法の場合、回答者の識別が開示の前提条件となるため、回答者の識別が不可能であれば、機密保護が保証される。
Sensitive cell: Cell for which knowledge of the value would permit an unduly accurate estimate of the contribution of an individual respondent. Sensitive cells are identified by the application of a dominance rule such as the (n,k) rule or the (p,q) rule to their microdata.	**センシティブなセル：**その数値を知ることで、個々の回答者の寄与を好ましくないほど正確に推定できるセル。センシティブなセルは、ミクロデータにおける(n,k)ルールあるいは(p,q)ルールのような占有度ルールの適用によって識別される。
Sensitive variables: Variables contained in a data record apart from the key variables, that belong to the private domain of respondents who would not like them to be disclosed. There is no exact definition given for what a 'sensitive variable' is and therefore, the division into key and sensitive variables is somehow arbitrary. Some	**センシティブな変数：**キー変数以外でデータレコードに含まれている変数。これらは、その開示を望まない回答者のプライベートに属している。' センシティブな変数 ' についての厳密な定義はなく、キー変数とセンシティブな変数との区別はある程度恣意的である。犯罪歴や病歴、債務記録など明らかにセ

data are clearly sensitive such as the possession of a criminal record, one's medical condition or credit record, but there are other cases where the distinction depends on the circumstances, e.g. the income of a person might be regarded as a sensitive variable in some countries and as quasi-identifier in others, or in some societies the religion of an individual might count as a key and a sensitive variable at the same time. All variables that contain one or more sensitive categories, i.e. categories that contain sensitive information about an individual or enterprise, are called sensitive variables.	ンシティブなデータがある一方で、例えば、個人の所得はセンシティブな変数とみなされる国や、準識別子とみなされる国もあり、また社会によって、個人の信仰している宗教が、キー変数とセンシティブな変数の両方に該当するなど、状況によって異なる。一つあるいは複数のセンシティブな区分、すなわち、個人あるいは企業についてのセンシティブな情報となる区分を含む全ての変数は、センシティブな変数と呼ばれる。
Shuttle algorithm: A method for finding lower and upper cell bounds by iterating through dependencies between cell counts. There exist many dependencies between individual counts and aggregations of counts in contingency tables. Where not all individual counts are known, but some aggregated counts are known, the dependencies can be used to make inferences about the missing counts. The Shuttle algorithm constructs a specific subset of the many possible dependencies and recursively iterates through them in order to find bounds on missing counts. As many dependencies will involve unknown counts, the dependencies need to be expressed in terms of inequalities involving lower and upper bounds, rather than simple equalities. The algorithm ends when a complete iteration fails to tighten the bounds on any cell counts.	**シャトル・アルゴリズム：**セル間の従属関係に基づいて繰り返すことで、上下のセル境界値を探る手法。クロス集計表では、個別の値と合計値の間に強力な従属関係が存在する。すべての個別の値が既知でなくても、ある合計値が明らかであれば、従属関係を用いることで、秘匿された値を推測することができる。シャトル・アルゴリズムでは、秘匿された値の境界を見つけるために、特定のサブセットを再帰的に繰り返すことにより、合計値と個別の値の従属関係を構築する。値の探索には、多くの従属関係が係わるので、従属関係は、単純な等式ではなく上限値、下限値と関連した不等式で表されなければならない。どのセル値についても、境界値が狭く繰り返せなくなると、アルゴリズムが収束する。
Special uniques analysis: A method of analysing the per-record risk of **microdata**.	**特別一意分析：**ミクロデータのレコードごとのリスクを分析する手法
Statistical confidentiality: The protection of data that relate to single statistical units and are obtained directly for statistical purposes or indirectly from administrative or other sources against any breach of the right to confidentiality. It implies the prevention of unlawful disclosure.	**統計的機密保護：**統計目的で直接入手、若しくは、間接的に行政機関、あるいは、他のソースから入手した客体に関するデータを、機密保持のために保護すること。これは、違法な開示を防止することである。
Statistical Data Protection (SDP): Statistical Data Protection is a more general concept which takes into account all steps of production. SDP is multidisciplinary and draws on computer science (data security), statistics and operations research.	**統計データ保護（SDP）：**統計データ保護は、統計作成の段階すべてを対象とした、より一般的な概念である。SDP は学際的で、コンピューター・サイエンス（データ・セキュリティ）、統計学及びオペレーションズ・リサーチを利用する。
Statistical disclosure: Statistical disclosure is said to take place if the dissemination of a statistic enables the external user of the data to obtain a better estimate for a confidential piece of information than would be possible without it.	**統計的開示：**統計の公開によって、データの外部ユーザーが、機密情報について確度の高い推定値を得ることが可能になった場合、統計的開示が起きたと言われる。

Statistical Disclosure Control (SDC): Statistical Disclosure Control techniques can be defined as the set of methods to reduce the risk of disclosing information on individuals, businesses or other organisations. Such methods are only related to the dissemination step and are usually based on restricting the amount of or modifying the data released.	**統計的開示抑制（SDC）：**統計的開示抑制は、個人、企業あるいはその他の機関に関する情報が開示されるリスクを抑制するための方法と定義することができる。この種の方法は公表段階のみに関係し、通常、公開するデータ情報量を制限するかあるいは修正する方法である。
Statistical Disclosure Limitation (SDL): Synonym of Statistical Disclosure Control.	**統計的開示制限：**統計的開示抑制（SDC）と同義
Subadditivity: One of the properties of the (n,k) rule or (p,q) rule that assists in the search for complementary cells. The property means that the sensitivity of a union of disjoint cells cannot be greater than the sum of the cells' individual sensitivities (triangle inequality). Subadditivity is an important property because it means that aggregates of cells that are not sensitive are not sensitive either and do not need to be tested.	**劣加法性：**(n,k)ルールあるいは(p,q)ルールの特性の一つで、補足的セルの探索に役立つ特性のひとつ。この特性は、共通の要素を持たずセンシティブでないセルを統合しても、セルの個々のセンシティブさの合計より大きくはならないことを意味する（三角不等式）。劣加法性は、センシティブでないセルの合計は、センシティブではなく、検証の必要がないことを意味することから、重要な特性である。
Subtraction: The principle whereby an intruder may attack a table of population counts by removing known individuals from the table. If this leads to the presence of certain zeroes in the table then that table is vulnerable to **attribute disclosure**.	**減法性：**統計表から既知の個別の数値を取り除くことにより、侵入者が、母集団の統計表に攻撃を仕掛けることができる原理。このことにより、統計表内の確実なゼロの存在が明らかになる場合、その統計表は、属性開示に対して無防備となる。
Suppression: One of the most commonly used ways of protecting sensitive cells in a table is via suppression. It is obvious that in a row or column with a suppressed sensitive cell, at least one additional cell must be suppressed, or the value in the sensitive cell could be calculated exactly by **subtraction** from the marginal total. For this reason, certain other cells must also be suppressed. These are referred to as **secondary suppressions**. While it is possible to select cells for secondary suppression manually, it is difficult to guarantee that the result provides adequate protection.	**秘匿：**統計表内のセンシティブなセルを保護する方法として最も広く利用されている方法の一つが秘匿である。秘匿されたセンシティブなセルを含む横列、若しくは、縦列においては、少なくともう一つのセルを秘匿しなければならない。そうでなければ、周辺和から差し引くことで、センシティブなセルの正確な数値が計算できてしまうことは明らかである。この理由から、他のセルの一部についても秘匿する必要がある。これらは、二次秘匿と呼ばれる。二次秘匿を行なうセルを手作業で選ぶことは可能であるが、適切な保護が講じられていることを保証するのは困難である。
SUDA: A software system for conducting analyses on population uniques and special sample uniques. The **special uniques analysis** method implemented in SUDA for measuring and assessing disclosure risk is based on resampling methods and used by the ONS.	**SUDA：**母集団一意及び特別な標本一意に関する分析を行うソフトウェア・システム。特別一意分析の手法は、リサンプリングの手法を基礎とし、英国国家統計局（ONS: Office for National Statistics）で用いられている開示リスクを測定、評価するための SUDA に実装されている。
Swapping (or switching): Swapping (or switching) involves selecting a sample of the records, finding a match in the data base on a set of	**スワッピング（またはスウィッチング）：**スワッピング（またはスウィッチング）は、レコードのサンプルを選定し、予め決められた

predetermined variables and swapping all or some of the other variables between the matched records. Swapping (or switching) was illustrated as part of the confidentiality edit for tables of frequency data.	変数についてデータベースの中で一致するレコードを見付け出し、その一致したその他の全ての変数か、あるいは、いくつかの変数を交換する処理である。スワッピング（またはスウィッチング）は、度数表作成のための機密保持エディットの一種である。
Synthetic data: An approach to confidentiality where instead of disseminating real data, synthetic data that have been generated from one or more population models are released.	**合成データ：**機密保持へのアプローチとして実際のデータを公開する代わりに、一つ、あるいは、複数の母集団モデルから発生させた合成データを公開すること。
Synthetic substitution: See **Controlled Tabular Adjustment**.	**合成データの置換：**Controlled Tabular Adjustment を参照のこと。
T	**T**
Table server: A form of remote data laboratory designed to release safe tables.	**集計表サーバー：**安全な集計表を公開するために設計されたリモート・データ・ラボラトリーの形式。
Tables of frequency (count) data: These tables present the number of units of analysis in a cell. When data are from a sample, the cells may contain weighted counts, where weights are used to bring sample results to the population levels. Frequencies may also be represented as percentages.	**度数表：**これらの表では、各セルの中に客体の数が表示されている。データが標本データであるとき、セルは、ウェイト付度数で構成されている。そのウェイトは、標本調査の結果を母集団レベルまで引き上げるために用いられる。度数は、また構成比として示されることもある。
Tables of magnitude data: Tables of magnitude data present the aggregate of a "quantity of interest" over all units of analysis in the cell. When data are from a sample, the cells may contain weighted aggregates, where quantities are multiplied by units' weights to bring sample results up to population levels.The data may be presented as averages by dividing the aggregates by the number of units in their cells.	**数量表：**数量表は、各セルの中に客体の " 当該数量 " の合計を示したものである。データが標本データであるとき、セルはウェイト付合計で構成されている。その分量は、標本調査の結果を母集団レベルまで引き上げるために客体のウェイトで掛け合わせる。各セルの中の客体の数で割ることにより、平均値として示されることもある。
Tabular data: Aggregate information on entities presented in tables.	**集計表データ：**表内に示された客体の回答を合計した情報
Target dataset: An anonymised dataset in which an intruder attempts to identify particular population units.	**ターゲット・データセット：**侵入者が、特定の母集団ユニットを識別しようとする匿名化データセット。
Threshold rule: Usually, with the threshold rule, a cell in a table of frequencies is defined to be sensitive if the number of respondents is less than some specified number. Some agencies require at least five respondents in a cell, others require three. When thresholds are not respected, an agency may restructure tables and combine categories or use cell suppression, rounding or the confidentiality edit, or provide other additional protection in order to satisfy the rule.	**閾値ルール：**通常、閾値ルールでは回答者の数が所定数を下回る度数表のセルは、機密セルに指定される。機関によって回答者数の下限を 5 としているところもあるが、他の機関は 3 としている。閾値を考慮しない場合、統計機関は、表を再構成し区分を統合したり、あるいはセルの秘匿、丸め法、機密エディット、ルールを満足させるために他の追加的な保護を提供する場合がある。

Top and bottom coding: It consists in setting top-codes or bottom-codes on quantitative variables. A top-code for a variable is an upper limit on all published values of that variable. Any value greater than this upper limit is replaced by the upper limit or is not published on the microdata file at all. Similarly, a bottom-code is a lower limit on all published values for a variable. Different limits may be used for different quantitative variables, or for different subpopulations.	**トップ及びボトム・コーディング：**これは、量的変数についてトップ・コードあるいはボトム・コードを設定するものである。ある変数のトップ・コードとは、その変数に関する全ての公表値の上限である。この上限を上回る数値はすべて、この上限値に置き換えられるか、あるいは公開されるミクロデータ・ファイルから完全に削除される。同様に、ボトム・コードは、ある変数に関する全ての公表値の下限である。 量的変数あるいは異なる部分母集団に対して、それぞれ異なる限界値が用いられる。
U	**U**
Union unique: A sample unique that is also population unique. The proportion of sample uniques that are union uniques is one measure of file level disclosure risk.	**共通一意：**母集団一意であり、また標本一意であること。共通一意であり標本一意でもある割合は、ファイルレベルの開示リスクの一尺度である。
Uniqueness: The term is used to characterise the situation where an individual can be distinguished from all other members in a population or sample in terms of information available on microdata records (or within a given key). The existence of uniqueness is determined by the size of the population or sample and the degree to which it is segmented by geographic information and the number and detail of characteristics provided for each unit in the dataset (or within the key).	**一意性：**この用語は、ミクロデータ・レコード（あるいは所与のキー）から得られる情報から、母集団や標本の中で、個人を他の母集団、あるいは、標本構成員と区別できる状況を特徴付けるために使われる。一意性が存在するかどうかは、母集団、あるいは、標本規模及び地理的情報によりどの程度セグメント化されているか、さらに、データセットにある各客体の特性の数及び特性の詳細さによって決まる。
Upper bound: The highest possible value of a cell in a table of frequency counts where the cell value has been perturbed or suppressed.	**上限値：**セル値が撹乱されたり、あるいは、秘匿された度数表におけるセル値の取り得る最大値。
V	**V**
Virtual safe setting: Synonym of remote data laboratory.	**仮想的安全環境：**リモート・データ・ラボラトリーと同義
W	**W**
Waiver approach: Instead of suppressing tabular data, some agencies ask respondents for permission to publish cells even though doing so may cause these respondents' sensitive information to be estimated accurately. This is referred to as the waiver approach. Waivers are signed records of the respondents' granting permission to publish such cells. This method is most useful with small surveys or sets of tables involving only a few cases of dominance, where only a few waivers are needed. Of course, respondents must believe that their data are not particularly sensitive before they will sign waivers.	**免責アプローチ：**集計表データを、秘匿する代わりに、ある統計機関では、回答者に対し、開示することで、その回答者のセンシティブな情報が正確に推計される原因になってしまっても公表を認めるよう要請することがある。これを、免責アプローチという。そのようなセルの公表を許可する回答者の署名入りの記録をとる。この方法は、小規模の調査や占有度の高い少数の事例が関わる表で、免責の必要件数が少ない時に有効な方法である。もちろん、回答者は、免責に署名する前に、そのデータが特にセンシティブでないことを確信する必要がある。

X	**X**
Y	**Y**
Z	**Z**

改訂版作成に当たり、和訳の変更を行った見出し語一覧

Exact disclosure	正確な開示	→	完全開示
Global recoding	グローバル・リコーディング	→	大域的再符号化（再格付け）
Local suppression	ローカル・サプレッション	→	局所秘匿
Perturbation based disclosure control methods	撹乱ベースの開示抑制法	→	撹乱的開示抑制法
Restriction based disclosure control methods	制限による開示抑制方法	→	制限的開示抑制法
Sensitive cell	デリケートなセル	→	センシティブなセル
Sensitive variables	デリケートな変数	→	センシティブな変数
Subadditivity	副加法性	→	劣加法性