

Evaluation Criteria for the Selection of a SDC Method

Johannes Rohde, Christiane Seifert (IT.NRW), Sarah Giessing (Destatis) (Germany)

Geheimhaltung-Statistik@it.nrw.de, kerstin.schmidtke@it.nrw.de, sarah.giessing@destatis.de

Abstract and Paper

In order to provide support during the selection process for a statistical disclosure control (SDC) method, the German National and State Statistical Institutes have developed a catalogue of evaluation criteria. It defines twenty criteria which a SDC method should ideally fulfill and assesses whether these criteria are satisfied by four frequently discussed methods in German Official Statistics. The article provides an introduction to the catalogue, presenting the criteria and discussing their interdependencies, elaborating especially on conflicting aims. All criteria are linked to the European Code of Practice and have been grouped into four main topics: protection against disclosure risks, data quality, accessibility and clarity, and cost effectiveness.

Evaluation Criteria for the Selection of a SDC Method

Johannes Rohde *, Christiane Seifert**, Sarah Giessing***

* Information und Technik Nordrhein-Westfalen (IT.NRW), Johannes.Rohde@it-nrw.de

** Information und Technik Nordrhein-Westfalen (IT.NRW), Christiane.Seifert@it-nrw.de

*** Statistisches Bundesamt (Destatis), Sarah.Giessing@destatis.de

Abstract: In order to provide support during the selection process for a statistical disclosure control (SDC) method, the German National and State Statistical Institutes have developed a catalogue of evaluation criteria. It defines twenty criteria which a SDC method should ideally fulfill and assesses whether these criteria are satisfied by four frequently discussed methods in German Official Statistics. The article provides an introduction to the catalogue, presenting the criteria and discussing their interdependencies, elaborating especially on conflicting aims. All criteria are linked to the European Code of Practice and have been grouped into four main topics: protection against disclosure risks, data quality, accessibility and clarity, and cost effectiveness.

1 Introduction

Statistical confidentiality represents a cornerstone of Official Statistics and is established as one of the principles defined in the European Statistics Code of Practice (CoP). The German federal law on Official Statistics rules the protection of personal data on a national frame and defines several exceptions to the general obligation to keep data relating to individual units confidential, such as the possibility to provide access to microdata for the purpose of academic research.

Due to the federal structure of Germany, consisting of a federal government and 16 federal state governments which almost all have an own Federal State Statistical Office, decisions regarding statistics published by the Federal Statistical Office as well as by the State Statistical Offices are brought out by management boards organized as working groups relating to the specific statistical domains. This includes the responsibility regarding the selection of a SDC method along with its implementation in consideration of specific requirements of the specialized statistics. In order to provide support to these domain specific working groups, the document “Evaluation Criteria for the Selection of a SDC method”¹ has been developed. This document roughly consists of two parts: First, twenty criteria relevant for the selection process of a SDC method are defined and described in detail. The second part evaluates four SDC methods recently discussed in German Official statistics (cell suppression, deterministic rounding, stochastic perturbation, and the micro aggregation method “SAFE”) along these criteria.

The present paper² provides a brief overview of the first part of the evaluation criteria document, i.e. the twenty selection criteria, which are to be considered during the

¹ The current version (Rohde et al., 2018a) is available at the joint intranet of the German Statistical Offices.

² For a German version of this paper see Rohde et al. (2018b)

selection process of a SDC method by the statistical domain working groups, are specified and explained. In this context, different aspects regarding practical, organizational, technical and legal issues are outlined and discussed from the point of view of statistical agencies as well as the relevant user groups. While keeping in mind that no SDC method is able to fulfill all requirements raised, the evaluation criteria document points out interdependencies of different criteria and observes possible conflicting aims. Since the document only provides arguments at a general level, no special cases of particular statistical domains are considered. On the same account, no prioritization of certain criteria is assumed.

The following main chapter of this paper presents a detailed description of the individual criteria, grouped into four main topics: protection against disclosure risks, data quality, accessibility and clarity, and cost effectiveness. Where applicable we point out when criteria can be directly derived from the European Statistics Code of Practice. The paper finishes with a short summary (section 3).

2 The Catalogue of Evaluation Criteria

2.1 Protection against Disclosure Risks

Principle 5 of the CoP claims confidentiality of all individual information of data providers and demands a corresponding legal provision. The German law on Official Statistics imposes a general obligation for confidentiality of individual data. Exceptions are only possible if aggregated data are published or individual information cannot be related to specific data providers. For this reason, a SDC method should ensure the protection of individual data against disclosure. The specific operationalization of the legal regulations in order to restrict disclosure risks depends on the selection of a SDC method and its parameterization.

2.1.1 Primary Protection of Confidentiality

The need for primary protection against disclosure risk arises from the following settings:

Identification, individual attribute disclosure: Rare or unique frequencies, in particular 1's and 2's, involve the possible identification of respondents and their corresponding attributes. This type of risk occurs if someone is able to use prior information in order to identify a respondent and derive information from publications of Official Statistics not known before. In the case of magnitude tabular data people with inside knowledge of e.g. a specific branch of economy, containing only a single or very few companies, may relate specific values and figures to a certain company, even if the frequency of companies behind these values is not published.

Group attribute disclosure: All respondents fall in just one or very few categories for a particular variable which may enable an intruder to identify an individual and its corresponding attributes, just by its affiliation to a group of respondents. This type of disclosure risk occurs when one cell of row or a column in a frequency table contains a

non-zero entry. Group attribute disclosure is also possible if one single respondent holds a unique attribute while all others contribute to only one different category. Group disclosure of higher order is possible if non-zeros only occur in specific categories which represent breakdowns of a category of higher order. E.g. if all residents of a community are either of Catholic or Protestant denomination, it can directly be derived that all residents are of Christian belief.

Approximate Disclosure: Magnitude tabular data involve the risk of close approximation of an individual value if a cell value is dominated by the contribution of one single or very few respondents. This type of risk typically occurs when providing statistics for branches of an economy with high concentration. Assuming insider knowledge for all respondents belonging to a certain branch, the magnitudes of the contributions from the respondents with the largest observations can be approximated by the difference of the cell value and the own contribution the other respondent as an upper bound. Special knowledge about particular attributes and its distribution may further reduce the upper bound, providing an even better estimate.

2.1.2 Secondary Protection of Confidentiality

A SDC method also must ensure protection against secondary disclosure risks. Implementation of secondary confidentiality protection guarantees sustainability of all provisions regarding primary protection (see 2.1.1). E.g., if a cell with primary risk would have been suppressed, without any further actions regarding secondary protection put into place disclosure would often still be possible: the suppressed cell could easily be uncovered by differencing the margin sum of the row or column and the sum of the corresponding unsuppressed cells. Typically, secondary disclosure risk arises from three different settings involving the calculation of differences: i) Geographical differencing, i.e. when information about a certain geographical and a superior entity is published in different tables, e.g. a county and its associated municipalities, ii) Differencing between a group and a subgroup, i.e. frequencies for certain attributes are published for a group and one of its subgroups. Then, the information about the remainder (anything not part of the subgroup) can be derived by differencing information about the superior group and the subgroup, iii) Differencing between overlapping tables, i.e. two tables relating to the same population spanned by different classifications of the same attribute. Joining information from the two different tables may result in secondary disclosure risk. Therefore, the total protective effect of a SDC method is determined by simultaneously accounting for primary and secondary disclosure risks.

2.1.3 Harmonized Protection of Confidentiality

The SDC method to be selected should allow harmonized concepts for the protection of individual data. All actions taken to secure confidentiality must be harmonized and co-ordinated for the entire program of publications for a given statistical domain, i.e. the SDC concept should not only include the protection of tables, but also map charts,

academic research output, publications in regional databases etc. Moreover, within a National (or European) statistical system an integrative SDC concept needs to be coordinated throughout all Statistical Offices involved in this system, in respect of temporal and organizational issues regarding publications. This aspect is closely related to criterion 2.1.2 as secondary disclosure risks arise as a direct consequence of uncoordinated protection of confidentiality. The all-over employment of the same SDC method for all kinds of publications also comprises the use of the same parameters.

2.1.4 Confidentiality in the Case of Third Party Access

This criterion expands the above defined criterion of harmonized protection (see 2.1.3) by another aspect. If legal regulations enable third parties outside of Official Statistics to request transmission of certain individual data, the protection against disclosure of individual attributes has to be maintained. The German law on Official Statistics allows certain authorities (administrations and ministries) to request original tables without SDC for the purpose of planning. In some cases also communal authorities with separate statistical units may demand individual data and publish their own statistical analysis, possibly on the basis of additional external data. In this case, the same requirements regarding confidentiality as imposed for Official Statistics are legally imposed for these authorities too. For this reason, the harmonized SDC concept of the statistical domain in question needs to be applied to publications of those third parties as well, including the use of the same parameters. In order to implement this, external authorities may need support (e.g. of the Statistical Institute) in order to be able to use the selected SDC method, i.e. SDC tools may have to be handed over to third parties with access to the data, training courses have to be designed and conducted etc. Taking these aspects into consideration, the costs and effort for the implementation of a new SDC concept may also increase (see 2.4.1 to 2.4.3).

2.2 Data Quality

Statistical publications have to comply with high quality standards as defined by the CoP, notably Principles 11-14. The protection of data confidentiality, however, always induces a certain level of information loss, regardless which SDC method is employed, and a loss of quality comes along with it. Hence, requesting a high level of data protection and securing high data quality at once represent main conflicting aims in the selection process of a SDC method. In the following, six criteria on SDC methods representing different aspects of data quality are defined and discussed.

2.2.1 Relevance

Principle 11 of the CoP demands a statistical program which meets the needs of the different user groups. The employment of a SDC method should guarantee that the needs of the data users can be satisfied as far as possible while protection of confidentiality is kept. For instance, the analysis at a high classification depth, like

publications at the municipality level, should not be impeded by the use of certain SDC methods as this would induce a loss of relevance from some user's point of view. Especially publications at a small scale level involve small frequencies and, thus, a high relative loss of information due to the employment of a SDC method. However, a SDC method should keep its applicability even at a small scale level on the one hand and maintain a high level of data quality on the other hand. The relevance of a statistic can also be impaired if results at a higher (geographical) level have to be suppressed to provide secondary protection for results at lower (geographical) level (see 2.2.3). In addition, the aspect of usability is another issue related to the criterion of relevance: If the choice of parameters for an SDC method results in a large number of suppressed cells or a high level of perturbation, *resp.*, the relevance of a statistic may also be depleted (see 2.2.2). Moreover, the criterion of relevance is closely related to the degree of complexity and transparency of SDC methods (see 2.3.1) as a nontransparent SDC process may also diminish relevance of a statistical publication or analysis.

2.2.2 Accuracy

The CoP sets the postulation that statistics should portray reality in an accurate and reliable manner (see Principle 12 CoP). This entails the requirement that the information loss induced by the employment of a SDC method should be kept at a minimum. Therefore, the loss of accuracy (along with the degree of protection) should be systematically examined by comparing different SDC methods and different parameter choices by means of appropriate key figures (such as mean absolute deviation or suitable distance measures³). If the employment of the SDC method results in heavy distortions of the univariate distribution of a single variable or the joint distribution of two or more variables, a high risk of drawing wrong conclusions from the output of Official Statistics would emerge. Depending on the SDC method and its parameterization, large deviations between the original and the published values can occur randomly or systematically. If systematic distortions are inherent to the SDC method (e. g. deterministic rounding) the distortion of margin values may increase when users sum up inner values of a table on their own. Limitations of reliability of certain values should be obvious for data users and clearly labeled in the publications (see 2.3.1). A high loss of accuracy may also lead to a loss of relevance for certain groups of users (see 2.2.1). If legal consequences directly arise from an output published by Official Statistics (e. g. Census populations) the SDC method should allow the publication of the original values in question without creating disclosure risks for other individual information.

2.2.3 Timeliness

Principle 13 of the CoP demands timeliness and punctuality of statistical publications. A loss of timeliness due to the employment of a SDC method should be avoided by a

³ A list of possible key figures for the examination of information loss can be found in the Catalogue of Criteria (Rohde et al, 2018).

far-sighted scheduling as far as possible. This aim can best be achieved if the SDC method allows the implementation of a coordinated and integrative SDC concept and, at the same time, if different statistical institutes are involved, a temporally independent application of SDC in those institutes. This is very relevant in the German Statistical System, if Federal State Statistical Offices release publications for a statistical domain before the Federal Statistical Office publishes aggregate results. On the other hand, a loss of timeliness may take place for several Statistical Institutes, if a centralized SDC process cannot be avoided in order to secure an effective and coordinated secondary disclosure protection. A possible automation of the entire SDC process, however, might reduce delays and may even result in time advantages in comparison to manually operated SDC processes (see 2.4.1).

2.2.4 Comparability

The necessity for coherence of statistical outputs is pointed out in Principle 14 of the CoP. It should be possible to combine, compare and reconcile statistics coming from related but different data sources. This possibility should not be impaired by the conduction of a SDC method. If a statistic builds on results of another statistic, then the protection of confidentiality of the original statistic may not be abolished. The criterion of comparability also includes the aspect of intertemporal comparability of statistics: A SDC method should allow the comparison of statistical results from different periods as well as the calculation of index numbers (such as rates of change). The conduction of the SDC method should avoid data gaps as well as large shifts of index numbers and heavy distortions or reversions of trends (see 2.2.2). Since every change of a SDC method over time produces a de facto methodical break which impairs comparability, the SDC method should only rarely be changed for a given statistic. In any case, the impacts resulting from the change of a SDC method are to be made transparent to the users (see 2.2.1, 2.3.1).

2.2.5 Additivity

Likewise, Principle 14 of the CoP covers the postulation that statistical tables should offer additivity properties. This criterion refers to the inner consistency of a table and is therefore closely related to criteria 2.2.4 and 2.2.6. A lack of additivity occurs if the SDC of inner values is conducted independently from the SDC of the margin values of a table. This results in the effect that the sum of the inner values of a row or a column is not equivalent to the corresponding margin value. As this may lead to confusions at first sight on the user's part, a clear and transparent communication of the SDC method along with all its impacts is most important when the SDC method may lead to non-additive results (see 2.3.1). These efforts may include the information that deriving aggregate values from the inner table values may result in unnecessary loss of information (see 2.2.2). In ideal circumstances, this kind of information loss can be avoided, if users are enabled to produce their own custom-designed output via a web

interface to a statistical database in which the entire SDC process automatically proceeds (see 2.3.5).

2.2.6 Consistency

Next to the criterion of comparability, as defined in 2.2.4, Principle 14 (especially Indicator 14.1) of the CoP claims consistency of tables, i.e. logically identical cells presented in different tables should observe the same value. A SDC method should therefore secure the outer consistency of statistical outputs, if cells of two tables refer to logically identical observations. If consistency cannot be maintained by the use of a SDC method, this may result in confusions on the users. Inconsistency may lead to multiple official results for one and the same issue and may hamper interpretations. Moreover, the entire statistical output may become attackable. Just like non-additivity, SDC methods which lack consistency especially demand clear and transparent information and reasoning regarding this effect and its impacts (see 2.3.1).

2.3 Accessibility and Clarity

In order to meet the goals set by Principle 15 of the CoP, statistics shall be presented in a clear and understandable form. Moreover, an easy and comfortable access to all statistical products must be provided to the users and the process of development of the statistics has to be transparently explained. The following six Criteria regarding aspects of accessibility and clarity refer to different indicators presented in Principle 15 of the CoP.

2.3.1 Complexity and Transparency

Indicator 15.1 of the CoP demands the presentation of statistics and corresponding metadata to facilitate a straightforward interpretation and meaningful comparisons. These requirements have to be maintained when a SDC method is employed. This leads to several properties, closely linked to each other, which ought to be fulfilled for the SDC method: First, users must be able to rely on a comprehensible explanation of how the SDC method works and which impacts on the data arise from its employment. For this reason, documentations have to be prepared which present the methodology and implications (i.e. advantages and disadvantages) of the SDC method for different user groups (occasional users, researchers, journalists etc.), as proposed by Indicator 15.6. This enables the users to come to accurate interpretations and draw the right conclusions from the statistical results. Data inaccuracy induced by the SDC method should be communicated to the users openly and - where appropriate- should clearly be labeled in the output, especially when a perturbative SDC method is employed (see 2.2.2). The degree of complexity of a SDC method largely determines to what extent the method can transparently be presented to the users. Highly complex methods are more difficult to explain than methods which are very straightforward or well known to users as well as within Statistical Institutes. Hence, the complexity of a SDC method affects the risk that corresponding documentations may not be understandable

to users and, consequently, that the entire SDC method may not be accepted. In the worst case, even the quality of statistical output could be mistrusted. In any case, the reason for the loss of information induced by the SDC method (as well as other impacts like non additivity) should be communicated. However, it has to be communicated in way that does not enable users to use this information to reduce the protection of individual data or uncover parameters.

2.3.2 Replicability

A well-known standard for academic research on the basis of empirical data is the requirement of replicability of statistical output. Analyses which are reproduced by using the same data basis and the same methodology are obliged to come to the same result. Hence, replicability is of significant importance for the reliability and credibility of academic research (see 2.3.6) as well as for the Official Statistics. A SDC method should therefore guarantee replicability of analyses using the same data basis and methodology.

2.3.3 Scope of Application

In some statistical domains mainly frequency counts data, in others also (or, mainly) magnitude data have to be protected against disclosure. At the beginning of the selection process, it must be clarified whether the SDC method is applicable for the data set in question. Considering the needs of the users, a SDC method should not only be applicable to statistical tables, but also to preferably many types of other statistical output, including multivariate analyses, grid data and time series. The application of SDC methods to procedures of multivariate statistics must ensure that the logic of the empirical correlations remains undistorted and computed measures and estimators (e.g. regression coefficients) are sustained according to the original data (see 2.3.6). Moreover, the comparability of time series must be guaranteed (see 2.2.4) and the possible generation of new attributes or categories when using a database via web access has to be taken into account (see 2.3.5). In order to uphold the integrative protection of disclosure, the SDC method should be applicable even to cartographic output presented in map charts (see 2.1.3). In all cases, the data quality standards as presented in section 2.2 should be fulfilled as far as possible. In addition, a SDC method has to be applicable when either two data sets of Official Statistics are matched with each other or when a data set of Official Statics is matched with external data (see 2.1.4 and 2.3.6).

2.3.4 Flexibility of Publications

Another aspect of accessibility is the preparation of customer-designed analyses to the public, as pointed out in Indicator 15.3 of the CoP. Usually, it is not predictable which combinations of statistical attributes are needed for standard publications, customer-designed analyses and research output. A SDC method should therefore be easily applicable, ideally carried out automatedly, even when an analysis or a table is not part

of a predetermined publication program. Otherwise, high effort would be necessary to (manually) check outputs for disclosure risks in order to account for both integrative and consistent confidentiality (see 2.1.3, 2.2.4). Since confidentiality of previous publications must be preserved, limitations of flexibility of publications could arise, if these publications are not part of the predetermined program. As a consequence, several data or analyses could not be offered to the public which would potentially result in a limited relevance for some user groups (see 2.2.1). A high degree of flexibility should also be guaranteed for the use within the statistical agencies. As pointed out in 2.2.3, a loss of timeliness could be prevented with SDC methods offering the option of a decentralized SDC process.

2.3.5 Web Access

The possibility of implementation of the entire SDC process into the structure of a database is closely related to the criterion of flexibility (see 2.3.4) since access to the statistics via databases could potentially enable users to produce custom-designed analyses. This criterion is inspired by Indicator 15.2 of the CoP which demands dissemination of the statistical products by modern information technologies, platforms and open data standards. Therefore, the SDC method should be capable of being integrated into the output production work flow of databases and this way allow the efficient generation of statistical tables and analyses that are safe for release. In order to maintain data confidentiality, close coordination between Statistical Offices is necessary (see 2.1.3). If results should directly be computed via web reporting, the SDC process must also be integrated within the computation process. In this case, the corresponding tools have to access the original data set which could potentially result in a higher implementation effort (see 2.4.1).

2.3.6 Data Use for Academic Purposes via Research Data Centers

The access to microdata for independent research purposes is stressed by indicator 15.4 of the CoP. In Germany, academic use of microdata is ensured by access possibilities offered via the Research Data Centers (RDC). Prior to their publication, all results computed by researchers are checked for possible disclosure risks. Since all rules and regulations of the SDC concept for a statistic have to be analogously applied by the RDC (see 2.1.4), several aspects which are most relevant to the RDC should be emphasized: Since users are exclusively responsible for the quality of their analyses, users rely on a high level of transparency of the SDC method and labeling of results which are of limited quality due to the employment of the SDC method (see 2.3.1). Moreover, replicability and consistency of results are indispensable to RDC users (see 2.2.6 and 2.3.2, *resp.*). A wide scope of applicability of a SDC method (see 2.3.3) is most essential for RDC users, since the legitimate freedom of research guarantees independent and self-reliant work with microdata to the researchers. Since different statistical procedures are applied to the data (regression analysis, factor analysis etc.)

and data sets are often matched with each other, it must be ensured that these potentialities are not impaired by the conduction of an SDC method.

2.4 Cost Effectiveness

The effective use of the available resources ensures cost effectiveness of statistical processes, as is pointed out by Principle 10 of the CoP. All efforts and potential implications resulting from the implementation of the SDC method into the work flow of a particular statistical domain should therefore be examined regarding their cost effectiveness. This section presents four criteria for the examination of cost effectiveness of a SDC method.

2.4.1 Implementation costs

The first step of the implementation process of a SDC method into the regular work flow of a particular statistical domain comprises effort for methodical planning and the elaboration of the SDC concept. During the conception stage, especially a sufficient level of time and human resources is needed. First of all, specific requirements regarding confidentiality, including the legal framework, must be clarified for the domain in question. By means of these requirements and with the support of the Catalogue of Criteria, the suitable SDC method can then be selected. If RDC offer access to the microdata of the statistics, the RDC are to be involved into the entire process of planning (see 2.3.6). Implementation costs during the conception stage largely depend on the selected SDC method: If the SDC method does not allow for flexible reporting appropriate to the user's demands (i.e. without need of co-ordinated output checks), the entire publication program along with the required levels of detail of the tables and analyses have to be predetermined (see 2.3.4, 2.3.5). Moreover, a time-consuming coordination process between Statistical Offices for the implementation of the SDC process could be inevitable. In a further step, the elaborated SDC concept must be checked and tested regarding its parameterization. This may result in increasing effort, if experts need to be consulted or if high computing time is necessary. After successful testing, the respective steps of the SDC process can be implemented into the work flow of the statistical domain, which may eventually require adjustment of existing routines. Algorithms and functions implementing the selected SDC method must be integrated into the standard production systems and codes or macros must be available for any other software used for output production, respectively. If corresponding IT solutions need to be developed beforehand, high implementation costs as well as substantial amounts of time and human resources may be required (see 2.4.3). In addition, training courses for the staff supposed to work with eventual new tools must be designed and time for their initial skill adaption should be scheduled. Suitable software tools and specific training courses must also be provided to the RDC, if access to microdata is offered for academic purposes.

2.4.2 Operating costs

As implied by Principles 10 and 13 of the CoP, a SDC method should allow the conduction of the SDC process and the subsequent approval for publication rapidly after the statistical results are checked (see 2.2.3). Thus, time and human resources should continuously be provided for the operational conduction of the SDC process. The feasibility of automatic integration of the SDC process largely determines type and amount of operating costs and should therefore be taken into consideration. If the SDC method can only be manually operated, the amount of tables to be checked and the complexity of the SDC method are of utmost importance. If an automatic implementation of the SDC method is possible, the degree of automation and the computing time are the relevant issues regarding operating costs. The computing time depends on the mathematical complexity of the SDC method and, potentially, on the performance of the employed hardware. If all steps of the production process can automatically be executed, decreasing costs for securing confidentiality in both the statistical domain departments and the RDC would cause increasing cost effectiveness. Just as for the conception stage, the possibility of flexible reporting and the preparation of analyses outside of the predetermined publication program are relevant for low operating costs resulting from the employment of the SDC method (see 2.3.4, 2.4.1). If results have to be provided in a large number of different aggregations, high computing time and a large amount of potential disclosure risks could arise with largely manually operated SDC methods. Finally, human resources must be provided for communication and consulting towards users of the statistical products. The SDC costs relevance of this issue depends on whether the SDC method is already well-known to the users and features a high degree of transparency (see 2.3.1).

2.4.3 Potential for standardized IT solutions

The standardization of SDC methods may entail increasing effectiveness of the work flow of the statistical processes. The CoP names the optimization of the productivity potential of information and communication technology for statistical processes and the implementation of standardized solutions as indicators for the effective use of resources (see Indicators 10.2 and 10.4 of the CoP). In order to even provide potential for the implementation of standardized software solutions, the separate steps comprising the SDC method should permanently be identical (see 2.4.2). If no standardized tools are available, possible savings resulting from the process standardization on one hand, and implementation costs on the other hand should be weighed against each other (see 2.4.1). The degree of automation (and to what extent manual actions are still necessary) also impacts the potential for standardized solutions. Furthermore, it should be checked whether IT solutions can directly be integrated into existing statistical software solutions (like SAS or R). When cost effectiveness is examined, a new software tool should not only be checked for its functionality, but also for its compatibility with technical and organizational issues of

the statistical agencies and their available amount of human resources as well as with the legal framework.

2.4.4 Usability of modern information and communication technologies on the data dissemination

Indicator 10.2 of the CoP demands an optimized potential of modern information and communication technologies not only at data processing (see 2.4.3), but also for the dissemination of statistical results. These technologies enable the users to generate flexible and largely customer-designed reporting. If the SDC method features these properties, no additional costs for time and human resources are required for the preparation of reports which were not included in the predetermined publication program. The degree of best possible exploitation of the potentials of modern technologies depends on the selected SDC method and whether a deposition of original data within the dissemination system is possible (considering the IT-security perspective). As distinct from criterion 2.3.4, this criterion focuses on the flexibility of a SDC method from the perspective of the statistical agency (and not from the user's point of view). It must be pointed out that this criterion is closely related to the criteria of relevance and accuracy (2.2.1 and 2.2.2, resp.) as well as to the other criteria of cost effectiveness defined in section 2.4.

3 Summary

This paper has introduced evaluation criteria relevant for the selection of a SDC method, classified into four main categories. The definitions and discussion presented in this paper represents a substantial part of the first chapter of the underlying German document by Rohde et al. (2019a) which cross-evaluates four SDC methods discussed frequently in the German Statistical System along the specified criteria. The entire document (comprising about 100 pages) passed the responsible management boards in February 2018. Since then, the evaluation criteria document constitutes the mandatory framework for the selection of a SDC method in German Official Statistics, supplemented by detailed descriptions of SDC methods, like e.g. in the Handbook on Statistical Disclosure Control (Hundepool et al., 2012).

References

- European Statistics Code of Practice for the National Statistical Authorities and Eurostat (EU statistical authority), adopted by the European Statistical System Committee on November 16, 2017. Available at: <http://ec.europa.eu/eurostat>
- Hundepool, Anco; Domingo-Ferrer, Josep; Franconi, Luisa; Giessing, Sarah; Schulte Nordholt, Eric; Spicer, Keith and de Wolf, Peter-Paul (2012). *Statistical Disclosure Control*. Chichester, UK: John Wiley & Sons Ltd.

Rohde, Johannes; Seifert, Christiane; Giessing, Sarah and Setzer, Stefanie (2018a) in cooperation with Jörg Breitenfeld, Stefan Brings, Jörg Höhne, Julia Höninger, Patrick Rothe and Ulrike Schedding-Kleis. *Entscheidungskriterien für die Auswahl eines Geheimhaltungsverfahrens. Version 1.0 [German only]*. Internal document of the German Statistical Offices, available at the joint intranet of the German Statistical Offices.

Rohde, Johannes; Seifert, Christiane; Giessing, Sarah (2018b). *Entscheidungskriterien für die Auswahl eines Geheimhaltungsverfahrens*. *Wirtschaft und Statistik*, No. 3/2018, Statistisches Bundesamt, p. 90-104.