American Journal of Public Health Research, 2025, Vol. 13, No. 6, 257-262
Available online at http://pubs.sciepub.com/ajphr/13/6/1
Published by Science and Education Publishing
DOI:10.12691/ajphr-13-6-1

Anonymization of Georeferenced Public Health Microdata

Simon Cremer, Lydia Jehmlich, Rainer Lenz*

Cologne university of technology, arts and sciences, Cologne, Germany
*Corresponding author: rainer.lenz@th-koeln.de

Received October 07, 2025; Revised November 09, 2025; Accepted November 17, 2025

Abstract Whether for the use of targeted advertising measures or tracing the spatial spread of viruses such as the recent corona virus: georeferenced microdata can - depending on the attributes it is provided with - hold enormous added value for society, science and research. However, the desired information can often not be extracted despite the inherent analytical content. The reason for this is that access to personal georeferenced datasets is severely restricted, as these are subject to statutory data protection. One way out of this dilemma is to apply a suitable anonymization method that guarantees data protection without significantly reducing the analytical validity of data. Based on the EU INSPIRE directive, the statistical offices of the EU are successively implementing the georeferencing of their surveys. This paper discusses selected anonymization methods being most promising for anonymizing georeferenced health data for research purposes, as they offer scope for combinations or more specific adaptations in order to balance out the trade-off between privacy and analytical validity of georeferenced health microdata.

Keywords: *data anonymization, disclosure risk, health microdata, location privacy, spatial analysis*

Cite This Article: Simon Cremer, Lydia Jehmlich, and Rainer Lenz, "Anonymization of Georeferenced Public Health Microdata." *American Journal of Public Health Research*, vol. 13, no. 6 (2025): 257-262. doi: 10.12691/ajphr-13-6-1.

1. Introduction

Georeferenced microdata has extraordinary potential for research and teaching, public administration and business. Key questions about the future and sustainability of our society can only be answered with high-quality and accessible geodata. For these reasons, the German Council for Social and Economic Data, together with the Federal Agency for Cartography and Geodesy and the state surveying authorities, set up a working group on "Georeferencing of data" back in mid-2010. Since then, great efforts have been made not only in Germany but also in the European Union as a whole to improve the European spatial data infrastructure. Based on the EU INSPIRE directive, the national statistical offices of the EU are successively implementing the georeferencing of their surveys. However, the real value of georeferenced data lies in its combination with other analysis features. This goes hand in hand with an increased need to protect confidential individual information, which in practice can only be achieved through elaborate anonymization processes. With the rapid progress of computing technology, digitization, and the simultaneous increase in the availability and linkability of large data sets, the importance of anonymizing georeferenced microdata is increasing to the same extent in order to remain on an equal footing with potential data attackers when passing on high-quality data products.

Anonymization of health data is a major challenge due to the high degree of individuality of personal and household data. According to the European General Data Protection Regulation (GDPR) Art. 9, these data are considered particularly sensitive and worthy of protection.

Hence, the bulk of publications specifically on the anonymization of geodata deals with health data. To identify confidential information, a potential data attacker can use the geocoordinates of the registration addresses of patients, for example. Not only tables of numbers, but also characteristics such as image data (X-rays, CT scans), names of treating physicians or administered medication can contribute to the identification of individuals. What are called "health apps" nowadays collect data on the health status of their users. The rapid development of data availability and methodology, combined with growing computing power, is increasingly calling traditional methods of anonymization and statistical confidentiality into question. Many anonymization procedures are based on the scenario of the potential data attacker who has additional knowledge in the form of quasi-identifiers (i.e. common attributes of additional knowledge and target data) and can thus identify sensitive information on respondants [1]. When applying anonymization methods, a bicriteria optimization problem must be solved. On the one hand, the analysis potential of the data should be maximized, while on the other the data should be protected as well as possible. In practice, the problem is

usually transformed into a single-objective optimization problem with constraints: Minimizing the information loss while specifying a threshold for data privacy. Or vice versa: minimizing the risk of re-identification by specifying a catalog of feasible analyses. With the digital revolution and the new digital data and algorithms that come with it, many familiar questions are being posed anew today and many new questions have been added.

The current paper has been presented at the 56ème Journées de Statistique of the French Statistical Society in Marseille in June 2025 [2].

2. Statistical Disclosure Control

A fine regionalization of individual data harbors a great risk of unambiguous assignments. Medical practices and clinics, for example, generally store various master data on their patients, such as name, address, date of birth, gender, insurance status, insurance number and the corresponding health insurance company. In the course of data confidentiality, this data is initially pseudonymized (i.e. deletion of the direct identifiers name, address and date of birth). The high sensitivity of the data usually requires the restriction to high aggregation levels, in applied research with microdata sets outside of research data centers usually at the level of appropriate administrative districts. To address this problem, the development of realistic privacy metrics that can better quantify the likelihood of and loss from potential re-identification with respect to geospatial data is essential. In addition, record linkage methods are proposed in the literature for the real verification of data security in anonymized microdata, with which geocoordinates can be stored in individual records and used for distance calculation [3]. According to the GDPR natural persons are protected by regulating the processing of data relating to an identified or identifiable natural person. The aim of anonymization is to strip the data of its personal reference. A potential attacker must not have any means at their disposal that they would generally use to identify a data subject (recital 26 sentence 3 GDPR). Anonymization of data in such a way that the remaining available information does not allow any conclusions to be drawn about individual subjects (for instance persons, households or companies), but still has sufficient information potential, is a core concern of every data producer. This is a basic principle of European statistics and aims to guarantee the confidentiality of data while at the same time maintaining its usability for statistical purposes [4].

3. Anonymity of Georeferenced Data

The what is called *k*-anonymity provides information on the degree to which data sets can be re-identified by combining their quasi-identifiers. A dataset is defined to be *k*-anonymous if each sequence of values appears with at least *k* occurrences in the data. The concept of *k*-anonymity can also be applied to spatial datasets, which is then referred to as spatial *k*-anonymity, being the most widely used metric for measuring anonymity in the masking of sensitive geodata [5]. It basically describes the number of masked points that are closer or equally close to the original location as the masked original point itself. The measure is skeched in Figure 1, where four points are closer or equally close to the original point than its associated masked one. Thus, within spatial *k*-anonymity, at least *k*-1 masked points are closer to an original address than the associated masked point. As a result, the probability of a data attacker locating the original location is at most 1/k.

However, there is also criticism of *k*-anonymity. For example, it does not provide protection against homogeneity. The problem of homogeneity arises when the data within a group is too similar or identical. If all individuals in an observed area show the same disease, *k*-anonymity does not offer effective protection.

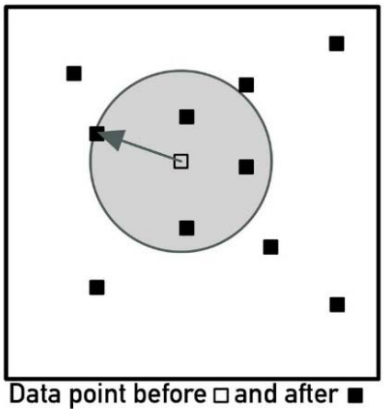

Figure 1. Spatial *k*-anonymity

An attacker possessing this information would still be aware, despite spatial *k*-anonymity, that every person in this area suffers from this disease. To counteract this, the concept of spatial *k*-anonymity can be extended with approaches such as l-diversity, *t*-closeness, or the concept of differential privacy [6].

4. Validity of Georeferenced Data

As already mentioned in the introduction, methods of data anonymization and statistical data confidentiality have to resolve a conflict of objectives. On the one hand, the information provided by the data subjects must be protected, on the other hand, procedures must be selected in such a way that the anonymized data still possesses sufficient analytical validity for the respective applications. Anonymization or the guarantee of confidentiality of individual data is always accompanied by a reduction of information.

Either information is suppressed to such an extent that it can no longer be assigned or only with disproportionate effort, or this protection is achieved by modifying the data to such an extent that the usefulness of it is (partially) lost. Either way, this means a loss of information. If a specific research objective is already being pursued with the data to be anonymized, the anonymization method can be specifically adapted to this objective. In the case of standardized data products, however, criteria must first be developed to preserve the analysis potential of georeferenced data sets. Typical questions regarding health data deal with the population's access to medical

care (specialists, hospitals, pharmacies, etc.), the classification of diagnoses (ICD-code) or even cultural facilities. Accordingly, there are many studies on the application of location masking methods to health datasets [7] such as leukemia [8], lung cancer [9] or, more recently, Covid-19 [10]. Nevertheless, further research is needed. [11] point out that "a substantive number of studies do not report uncertainty levels of the spatial data and its potential impact on their analytical results". This makes analyses based on these data susceptible to bias. Incorrectly drawn conclusions are "followed by wrong decision-making of local health policies which target specific geographical areas".

5. Preselected Methods of Anonymization

This chapter presents pre-selected anonymization methods that serve as basic mechanisms for anonymizing georeferenced microdata. Accordingly, these have proven to be particularly suitable for further studies on anonymization of health data, as these methods offer scope for combinations or more specific adaptations.

First of all, the considered dataset is pseudonymized, which is in most cases not the last step to protect confidential data. During pseudonymization, direct identifiers such as name or address are replaced by a pseudonym, for instance a sequential number or a code, to make it more difficult to correctly assign individuals. The further anonymization can then be roughly divided into two categories: aggregation and perturbation of quasi-identifiers and sensitive attributes.

The following sections present methods from both areas, which – in the opinion of the authors - are most promising for anonymizing health microdata. It should also be mentioned that these methods are in general likely to be combined in practice.

5.1. Aggregation of Data

In principle, a large number of prominent clustering methods can be used to aggregate geodata. Clustering can be carried out according to spatial, but also to non-spatial criteria such as age or gender. Spatial clusters can be formed, for example, w.r.t. administrative units or by grid lines [5]. Very promising are grid masking and areal aggregation. The latter is the most frequently used approach for masking georeferenced data, e.g. see [12]. For instance, regarding so-called disease mapping or in the presentation of morbid-mortality atlases [13].

Applying *grid masking*, all original data points are aligned to uniform grid cells [14]. The decision maker then selects an appropriate size of the cells. A regularly selected grid cell has a side length of 100m, 125m or 250m, see e.g. [14] or [10]. Applying *grid line masking*, the original points are moved to the nearest edge of the grid cell surrounding them. An actual aggregation occurs when several points have the same nearest grid line and are combined into one single point. Applying *grid centroid masking*, all points in the interior of some cell are moved to the corresponding cell centroid.

Instead of a large number of data points being reflected by one single point, the surrounding area can represent a cloud of data points. An interesting contribution treating Adap-tive Areal Elimination (AAE) and a continuation of the Adaptive Areal Masking (AAM) is found in [15]. Although aggregation methods generally promise a high level of data protection, the original data is heavily altered, resulting in a loss of spatial information. Another major task is the modifiable areal unit problem (MAUP), describing the fact that aggregated data values always depend on the specific setting of boundaries. As we see in the example in Figure 2, boundaries are decisive in the evaluation of aggregation values.

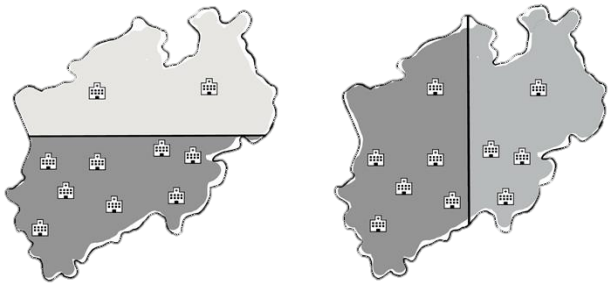

Figure 2. MAUP using the example of hospital density

In Figure 2 on the left, there is a low (light gray) and high density (dark gray) of hospitals with horizontal boundaries. If the boundary is drawn vertically (right example), there is a medium density (medium gray) in addition to a high density. In terms of hospital construction, less need for action could be interpreted in the case on the right if the original data is not available.

An administrative boundary does not usually represent a barrier for individuals in terms of access to health-related services in everyday life [16]. For research purposes, however, the situation is different. For example, a correct assessment of medical care density requires supra-regional data and not aggregations of administrative units. This further complicates the identification of regional clusters. In addition, the presentation as a choropleth map can not only hinder interpretations but also encourage misinterpretations. Thus, the uniform coloring of a surface leads to the false assumption that the variable of interest is evenly distributed [17]. The representation of values in the form of different intensity coloring means that attention is often drawn to the dark areas [18]. Another challenge is the dynamic nature of health data. Health data has the potential to quickly become outdated. The use of data over long periods of time, for example, leads to a bias in estimates of morbidity rates. The excess found in some areas may therefore only reflect a past situation that still exists due to the aggregation of information [13].

5.2. Perturbation of Data

The most widely used approach for coding an original position is to apply a random perturbation [8,19,20]. In contrast to aggregative methods, the number of data points generally remains the same for both deterministic and stochastic perturbation.

Affine point transformations are the most basic geographic masking methods. In an affine transformation, the displacement of points is deterministic [21]. Such

masks, in which the number of points after masking corresponds to the number before masking, but the coordinates of the points change, are also called isomasks [22]. In general, it can be observed that this approach has not gained acceptance in the scientific community [23]. Affine point transformation shows the advantage that the relative position of points is maintained, i.e. it preserves spatial patterns. At the same time, a data attacker can obviously succeed in re-identifying a large part, if not the entire dataset, based on just two unmasked points. It can also be used sophisticated pattern recognition for identification of original data points. Therefore, masking by affine point transformation is not considered reliable. Furthermore, the concatenation of different affine transformations (which itself determines an affine transformation) can cause additional work without achieving any progress in terms of the anonymity/quality ratio. In general, it can be observed that this approach has not gained acceptance in the scientific community. For this reason, these methods will not be discussed in detail.

The following data perturbing methods are considered to be suitable for the anonymization of health data. It is possible and even recommended to use these methods adaptively depending on the population density.

Voronoi diagrams have been used in mathematics since the beginning of the 20th century. [14] were the first to apply the concept of *Voronoi masking* in the context of geodata. According to them, Voronoi polygons are placed around the points to be masked. Each point defines the center of gravity of its surrounding polygon. In addition, the polygons have the property that their boundary lines always run exactly in the middle of two points. After the polygons have been created, each point is placed on the nearest edge of its surrounding polygon (see Figure 3). Although Voronoi masking is assigned to the data perturbing methods, it obviously contains aggregative elements. This is the case when several points are snapped to the same point on the polygon boundary. In contrast to the aggregative methods, in case of Voronoi msking it is generally impossible for the data user to recognize the margins of the aggregates.

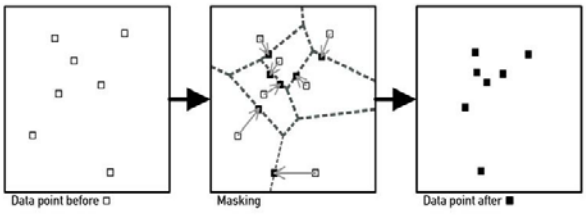

Figure 3. Voronoi masking

The advantage is that the displacement distance in densely populated areas is smaller than in sparsely populated ones. Thus, the pattern of anonymized points closely resembles the original distribution. If several points are close to each other in a remote region, these are also only moved a short distance. This can lead to smaller displacement distances than with concealment techniques that do not take such patterns into account [14]. Another advantage is that no relocated point is placed on another household, while it can be critically questioned whether Voronoi masking can be meaningful at all in a complete dataset, as the spatial *k*-anonymity could in some areas reduce to $k = 2$ (compare Figure 3). A larger k can then only be achieved if the Voronoi masking procedure is run several times. As mentioned in the beginning of section 5.2, adaptive variants of Voronoi masking (see [24]) should also be considered in future investigations.

The *donut masking* method is a continuation of the random distribution within a circular region [12,25]. In addition to a circular region with the radius r_{max}, into which some point is randomly displaced, an inner ring with the radius r_{min} is defined around the original location, beyond which the displacement must extend. Thus, two radii, which resemble the shape of a donut, determine the minimum and maximum distance between which the new coordinates of the point have to lie. If, in addition to the torus shape, there are further restrictions with regard to the new placement, so-called "eaten donuts" may result. In other words, donut shapes in which a piece is missing because this space is not defined for the new placement of the point. Examples of such restrictions are borders of any kind or uninhabitable places such as parks, bodies of water or mountains.

In addition, the donut mask can also be designed adaptively. Adaptive masks are masks whose anonymization parameters vary to meet specific anonymity requirements based on the underlying population density [12]. For the donut method, this specifically means that both the minimum and maximum distance are selected depending on the population density at the original location. If there is a high population density, radii can be selected correspondingly smaller than in rural areas.

The application of a suitable *Bimodal Gaussian* distribution to the variable X of geocoordinates is in some ways related to donut masking. Instead of the uniformly distributed probability, where the masked point is placed between some inner and the outer radius, a bimodal Gaussian distribution is used for the random distance function [10]. However, the basic idea to apply this bimodal distribution to continuous microdata has already been implemented successfully in the context of anonymizing business microdata [1]. The coordinates of the original location (x_i^{orig}, y_i^{orig}) form the coordinate origin. A distance d_i to this origin is chosen at random from a normal distribution with expectation μ and standard deviation σ, that is, $X \sim N(\mu;\sigma)$, resulting in masked coordinates (x_i^{mask}, y_i^{mask}). In practice, different values of μ and σ are tested and several calculation runs are carried out so that an optimum result can be achieved. The displacement angle θ_i is also selected randomly, namely from the uniform distribution $X \sim G(0;2\pi)$. Hence, it follows:

$$\left(x_i^{\text{mask}}, y_i^{\text{mask}}\right) = \left(x_i^{orig} + d_i \cos(\theta_i), y_i^{orig} + d_i \sin(\theta_i)\right)$$

Within the darker inner area of the torus (middle of Figure 4) the probability of placement is highest. The probability of placement decreases towards the outer and inner edges of the ring before it converges to zero (white area).

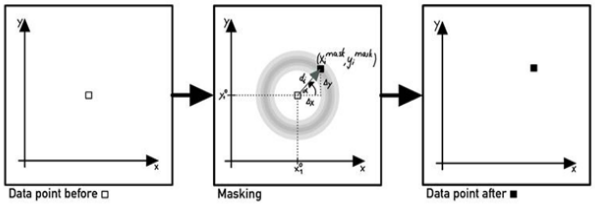

Figure 4. Bimodal Gaussian perturbation

Data swapping encompasses various approaches. The term "data swapping" goes back to the early work [26]. The authors developed the idea of publishing tabular data without being able to trace sensitive data back to individuals. In addition to swapping values within a characteristic or replacing one characteristic value with another, this approach can also be applied to geodata. Which involves swapping “records from one place to another, so that information from an individual with a certain set of key attributes is exchanged with the information from another individual, located in a different geographical area, but who matches the same attributes" (see [27], p.14). In the event of a data attack, the detected location data is no longer linked to some original personal data.

The *location swapping* method introduced by [28] begins with the creation of a buffer zone with a defined radius around the point to be moved. This radius is defined adaptively depending on the population density around this location. An address is then randomly selected from the available (real existing) addresses in that buffer zone. The point is then shifted to this address. [28] suggest the addition of elements of the donut method. An inner radius is defined for a buffer zone, which describes the space in which the point to be moved must not lie. The inner radius is determined in relation to the outer radius. Obviously, the setting of radii can also be carried out adaptively depending on the population density. The strength of data swapping obviously lies in maintaining the geographical distribution of points, while multivariate correlations can be affected depending on specific users’ interests.

The what is called *verified neighbor mask* works very similarly to location swapping. To carry out the method, the geographical coordinates of the original locations and a point data layer of publicly available property parcel centroids with information about residential status and other variables of interest are required [29]. The value of the desired k-anonymity must be selected before the application starts. The pool of neighbors is created based on the attributes of the parcel centroids. This means that proxy points (or addresses) are selected in areas with residences that are close to the original location and can plausibly replace it. If desired, the nearest neighbors of the original points can also be excluded from the pool formation [30]. This makes the masking similar to donut location swapping. Optionally, after this pool creation, another selection can be made according to administrative boundaries and environmental variables. Finally, a value is randomly selected out of the *k*-nearest neighbors for each original location, to which the original data is then transferred.

6. Conclusion and Future Work

As introduced at the beginning, current research indicates that the anonymization method should be selected depending on the research objective and the intended data product. Accordingly, no method can be named independently as better or worse. The methods of data aggregation and data perturbation presented have shown so far to be suitable for anonymizing georeferenced health microdata. They do have different strengths and weaknesses.

The aggregating methods are contrasted with the perturbation methods. The adaptation of masking parameters to the underlying population density is advantageous; as sparsely populated regions require stronger and more densely populated regions weaker anonymization measures to ensure anonymity. However, adaptive random shifts are often based on the assumption that the population is approximately homogeneously distributed. This condition is often not fulfilled, so that in very heterogeneously populated areas there may be lower actual k-anonymity than expected [31]. Our selection of advantages and drawbacks shows that there is a need for further research. It is therefore important to carry out empirical studies with real world geodata to compare the methods presented here. This is the only way to assess the suitability of individual masking methods for specific applications. Aiming to establish a ranking of the pre-selected methods and to provide suitable combinations of them for different datasets and ways of data access.

The overarching goal is to derive recommendations for application and tools to generate anonymized research data of highest possible analysis potential. A scientifically sound portfolio of methods must be developed in a discursive process involving all stakeholders in order to meet the strict data protection requirements on the one hand and the diverse user requirements on the other. Here, a distinction should be made between the possible ways of accessing data, including data dissemination as standardized scientific use files, as what are called campus-files for teaching purposes, at a guest scientist's workstation or by remote data access [32]. The research data centres of the Federal Statistical Office and the Statistical Offices of the Federal States in germany are in the process of gradually geocoding their data [33]. Moreover, the what is called Health Data Lab is responsible for making anonymized claims data from all statutory health insurance beneficiaries in Germany available to research [34].

Germany has transposed the EU INSPIRE directive into national law with the Geodata Utilization Ordinance and has committed itself to making those INSPIRE-relevant data available as open data. The draft Health Data Usage Act already aims to establish a decentralized health data infrastructure with a central data usage and coordination office. In particular, this should facilitate the usability of the data for public welfare purposes. The current research project *AnigeD*, funded by the German Ministry of Research, Technology and Space, is intended to lay the methodological foundation for this.

Conflict of Interest and Funding State-ment

The authors declare that they have no financial or non-financial conflicts of interest that could have influenced the work reported in this manuscript.

This work was carried out within the research cluster "Anonymization of integrated and georeferenced Data" (AnigeD), which is supported by the Research Network Anonymization for Secure Data Use of the German Ministry of Research, Technology and Space and by the Federal Government's research framework programme on IT security "Digital. Secure. Sovereign." funded by the European Union – NextGenerationEU.

References

[1] Ronning, G., Sturm, R., Höhne, J., Lenz, R., Rosemann, M., Scheffler, M. and Vorgrimler, D., *Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten*, Statistisches Bundesamt, Wiesbaden, Series 'Statistik und Wissenschaft', vol. 4, 2005.

[2] JdS 2025, 56ème Journées de Statistique de la Societé Francaise de Statistique, Université Aix-Marseille, june 2025, see: https://jds2025.sciencesconf.org/?lang=fr.

[3] Lenz, R., Measuring the disclosure protection of micro aggregated business microdata - an analysis taking as an example the German Structure of Costs Survey, *Journal of Official Statistics 22 (4)*, 681-710, 2006.

[4] Eurostat, Statistical confidentiality and personal data protection, 2023. Available at: https:// ec.europa.eu/ eurostat/ de/web/ microdata/statistical-confidentiality-and-personal-data-protection.

[5] Broen, K., Rob, T. and Jon, Z., Measuring the impact of spatial perturbations on the relationship between data privacy and validity of descriptive Statistics, 2021, *International Journal of Health Geographics*, 20 (3).

[6] Dwork, C., Differential privacy, *International colloquium on automata, languages, and programming*, 1-12, 2006.

[7] Gao, S., Rao, J., Liu, X., Kang, Y., Huang, Q., App, J., Exploring the effectiveness of geomasking techniques for protecting the geoprivacy of Twitter users, *Journal of spatial inform. science, 19*, 105–129, 2019.

[8] Armstrong, M.P., Rushton, G. and Zimmermann, D.L., Geographically Masking Health Data to preserve Confidentiality, *Statistics in Medicine*, 18, 497-525, 1999.

[9] Kwan, M., Casas, I. and Schmitz, B. (2004), Protection of Geoprivacy and Accuracy of Spa-tial Information: How Effective Are Geographical Masks? *Cartographica*, 39(2), 15-28, 2004.

[10] Houfaf-Khoufaf, W. and Touya, G., Geographically Masking Addresses to Study COVID-19 Clusters, *Univ. Gustave Eiffel*, 2021.

[11] Delmelle, E. M., Desjardins, M. R.; Jung, P., Owusu, C., Hohl, A., Dony, C., Uncertainty in geospatial health: challenges and oppurtunities ahead, *Annals of Epidemology, 65*, 15-30, 2022.

[12] Hampton, K.H., Fitch, M.K., Allshouse, W.B., Doherty, I.A., Gesink, D.C., Leone, P.A. and Miller, W.C., Mapping Health Data: Improved Privacy Protection with Donut Method Geomasking, *American Journal of Epidemiology*, 172 (9), 1062–1069, 2010.

[13] Ocaña-Riola, R., Common errors in disease mapping, *Geospatial Health*, 4 (2), 139-154, 2010.

[14] Seidl, D.E., Paulus, G., Jankowski, P., and Regenfelder, M., Spatial obfuscation methods for privacy protection of household-level data, *Applied Geography*, 63, 253-263, 2015.

[15] Kounadi, O. and Leitner, M., Adaptive areal elimination: A transparent way of disclo-sing protected spatial datasets, *Computers, Environment and Urban Systems*, 57, 59–67.

[16] Koller, D., Wohlrab, D., Sedlmeir, G. and Augustin, J., Geografische Ansätze in der Gesundheitsberichterstattung, *Bundesgesundheitsblatt*, 63, 1108–1117, 2020.

[17] Erfuhrt, K., Groß, M., Rendtel, U., Schmid, T., Kernel density smoothing of composite spatial data on administrative area level - A case study of voting data in Berlin, *AStA Wirtsch Sozialstat Arch, 16*, 25–49, 2022.

[18] MacEachren, AM., *How Maps Work: Representation, Visualization, and Design*, The Guilford Press, 1995.

[19] Zandbergen, P. A., Ensuring Confidentiality of Geocoded Health Data: Assessing Geographic Masking Strategies for Individual-Level Data, *Advances in Medicine*, 2014.

[20] Swanlund, D., Schuurman, N., Zandbergen, P., Brussoni, M., Street masking: a network-based geographic mask for easily protecting geoprivacy, *International Journal of Health Geographics, 19 (26)*, 1-11, 2020.

[21] Wang, J., Kim, J., Kwan, M.-P., An exploratory assessment of the effectiveness of geomasking methods on privacy protection and analytical accuracy for individual-level geospatial data, *Cartography and Geographic Information Society, 49 (5)*, 385-406, 2022.

[22] Kounadi, O., Towards geoprivacy guidelines for spatial data, *ETH Zürich Research Collection*, ETH Library, 2015.

[23] Leitner, M., Curtis, A., A first step towards a framework for presenting the location of confidential point data on maps—results of an empirical perceptual study, *International Journal of Geographical Information Science, 20 (7)*, 813-822. 2006.

[24] Polzin, F., Kounadi, O., Adaptive Voronoi Masking: A Method to Protect Confidential Discrete Spatial Data, *11th International Conference on Geographic Information Science*, 2021.

[25] Cremer, S., Jehmlich, L. and Lenz, R., Masking georeferenced health data - an analysis taking the example of partially synthetic data on sleep disorder, *Privacy in Statistical Databases, Domingo-Ferrer and M. Önen (Eds.)*, LNCS 14915, 297–309, 2024.

[26] Dalenius, T., Reiss, S. P., Data Swapping: A Technique for Disclosure Control, Journal of Statistical Planning and Inference, 6, 73-85, 1982.

[27] Gutmann, M. P., Witkowski, K., Colyer, C., McFarland O'Rourke, J., McNally, J., Providing Spatial Data for Secondary Analysis: Issues and Current Practices Relating to Confidentiality, *Population Research and Policy Review 27 (6)*, 639-665.

[28] Zhang, S., Freundschuh, S.M., Lenzer, K., Zandbergen, P.A., The Location Swapping Method for Geomasking, *Cartography and Geographic Information Science,*44 (1), 22-34, 2017.

[29] Richter, W., The verified neighbor approach to geoprivacy: An improved method for geographic masking, *Journal of Exposure Science and Environmental Epidemiology 28*, 109–118, 2018.

[30] Swanlund, D., Schuurman, N., Zandbergen, P., Brussoni, M., Street masking: a network-based geographic mask for easily protecting geoprivacy, International Journal of Health Geographics, 19 (26), 1-11, 2020.

[31] Li, N., Li, T. and Venkatasubramanian, S. (Eds.), t-closeness: Privacy beyond k-anonymity and l-diversity, *IEEE 23rd international conference on data engineering*.

[32] Lenz, R., On the way to remote access to German official microdata, Statistique et nouvelles technologies de l'information, *Revue des Nouvelles Technologies de l'information*, 125-138, 2011 (paper presented at the 41ème Journées de Statistique de la Societé Francaise de Statistique, Bordeaux 2009).

[33] Research data centres of the Federal Statistical Office and the Statistical Offices of the Federal States, https://www.forschungs-datenzentrum.de/de, last accessed on 3 september 2025.

[34] Wenzel, M., Corr, D., Riedel, N., Hapfelmeier, J., Zimmermann, L., *State of efficient research on secondary health data in the German health data lab*, Georg Thieme Verlag, Stuttgart, 2025.