# Response to the Public Consultation on Guidelines 02/2026 on Anonymisation

Comments on the draft Guidelines adopted by the European Data Protection Board on 7 July 2026

**Swiss Data Anonymization Competence Center**

Olten, October 2026

# Impressum

## Recommended citation

Thees, O., Müller, R., & Templ, M. (2026). *Response to the public consultation on Guidelines 02/2026 on Anonymisation* [Consultation response]. Swiss Data Anonymization Competence Center (SwissAnon), University of Applied Sciences and Arts Northwestern Switzerland FHNW.

## Project team

Oscar Thees (contact, oscar.thees@fhnw.ch)
Roman Müller
Prof. Dr. Matthias Templ

## Submission

Submitted to the European Data Protection Board.
Stakeholder category: academic / research institution. Country: Switzerland.
Consultation closes 30 October 2026.

Submitted in response to the public consultation opened by the European Data Protection Board. Contributions are published on the EDPB website.

Swiss Data Anonymization Competence Center (SwissAnon)
University of Applied Sciences and Arts Northwestern Switzerland FHNW, Olten, Switzerland
swissanon.ch

# 1 Introduction

**In one paragraph**

We welcome the draft Guidelines (European Data Protection Board [EDPB], 2026). They establish a workable framework for assessing anonymisation and will be of considerable value to practitioners in the field. However, we have two main comments. First, the criteria are not mapped to the established terms of identity, attribute and membership disclosure, so practitioners cannot readily draw on existing methods and measures. Second, the No Inference test in footnotes 43 and 44 asks whether a person's own record was needed to draw an inference about them, and treats the inference as knowledge about the population at large whenever it was not. In the draft's own Example 19, three women of the same year of birth and postcode all have lung cancer, and the draft rightly concludes that anyone who knows that their neighbour is one of them can infer her diagnosis. Yet the test in footnotes 43 and 44 would be passed, because the other two records show the same diagnosis, so her own record is not needed. We propose that the prior belief be based on people like her who are not in the released data. Further comments follow paragraph by paragraph in Section 3.

## 1.1 Who we are

The Swiss Data Anonymization Competence Center (SwissAnon, swissanon.ch) is a research centre at the University of Applied Sciences and Arts Northwestern Switzerland FHNW, led by Prof. Dr. Matthias Templ. We develop methods and open-source software for statistical disclosure control and synthetic data, including the R package `sdcMicro` (Templ, 2017; Templ et al., 2015), teach courses on anonymisation, and carry out anonymisation projects and audits for companies and public bodies. We currently work on the Swiss National Science Foundation Bridge Discovery project "Harnessing event and longitudinal data in industry and health sector through privacy preserving technologies" (grant 211751), on two projects for Nestlé and on one for Swiss Federal Railways (SBB). In both Nestlé projects, we were asked to anonymise data in accordance with the draft Guidelines, which has raised practical questions about how the criteria can be interpreted and evidenced in a reproducible assessment.

We respond as a research centre, not as a controller with an interest in a particular outcome. Our interest is that the framework works in practice and that its results can be checked. Where we consider a criterion to be underspecified, we indicate what evidence would be needed to demonstrate that it is met. Our comments are not exhaustive: they focus on the questions we encountered when applying the criteria to real data, and the absence of a comment on other parts of the draft should not be read as agreement. We would be glad to support the EDPB further if this would be useful.

## 1.2 Key points

- **Established terminology.** Map No Record Isolation, No Linkage and No Inference to identity, attribute and membership disclosure, so practitioners can use existing methods and measures (Section 2.1).
- **A No Inference test that catches disclosure in homogeneous groups.** Measure the prior belief against people who share the characteristics the entity knows but whose data is not necessarily in the released data, not against the remaining members of the same group. Where a group shares the same sensitive value, presume the inference is meaningful unless the controller shows otherwise (Section 2.2).
- **Evidence and reproducibility.** State what an assessment must document and which evidence is acceptable, including for synthetic data, AI models and longitudinal data (Section 3).

# 2 General remarks

## 2.1 The three criteria should be mapped to established disclosure risk terms

The draft assesses anonymisation against three criteria: No Record Isolation, No Linkage and No Inference. These terms do not match the terminology commonly used in the statistical disclosure control literature. The *Handbook on Statistical Disclosure Control* distinguishes two types of disclosure risk, *identity disclosure* and *attribute disclosure* (Hundepool et al., 2026, Section 1.1). The Federal Committee on Statistical Methodology (FCSM, 2005) adds *inferential disclosure*, which corresponds most closely to the draft's No Inference criterion. Especially in the context of synthetic data, *membership disclosure*, the inference that an individual's record is included in the data, is mentioned as well. Practitioners who prepare or review anonymisation assessments, including statistical offices, research data centres and providers of anonymisation tools, are usually trained in this terminology. The methods, risk measures and software they use are organised around it. The draft's criteria descend instead from the three risks set out by the Article 29 Working Party – singling out, linkability and inference (Article 29 Data Protection Working Party, 2014).

The correspondence is close but not one-to-one. No Record Isolation and No Linkage both address forms of identity disclosure: isolating the record of an individual within the given data, and linking it with other data relating to that individual. No Inference covers attribute disclosure, and the draft's discussion of whether an entity knows that an individual is included in the given data (paragraph 94, Example 22) touches on membership disclosure without naming it. The draft's own examples show the gap: Example 12, placed under the No Linkage criterion, shows that all patients with a given sex, year of birth and postcode share the same diagnosis. The diagnosis can be learned without knowing which of these records belongs to the patient, which is attribute disclosure. Without an explicit mapping, practitioners must reconstruct it themselves, and assessments may diverge depending on how each criterion is translated into familiar risk measures.

We propose that the Guidelines state how each criterion relates to identity, attribute, inferential, and membership disclosure, for example in a short table or footnote where the criteria are introduced, and refer to the established literature. This would make the criteria easier to apply with existing methods and would reduce the risk of confusion between the two vocabularies.

## 2.2 Attribute disclosure can be argued away as population knowledge

The No Inference criterion is meant to catch attribute disclosure: learning a sensitive fact about a person from the released data, even where their record cannot be singled out. Whether this has happened can be seen by comparing what an entity could believe about a person before and after seeing the data (Duncan & Lambert, 1986). Suppose you are such an entity: you know that your neighbour, a woman born in 1951 who lives in postcode area 32***, was treated at a research institute that treats patients with a wide range of conditions, and that the institute's released data covers all its patients. Without the released data, the best guess about whether she has lung cancer is how common that diagnosis is among *comparable people*: people who share what the entity knows about her (women born in 1951, from the same region, treated at such an institute) but whose data is not necessarily in the release. That share, perhaps 1%, is the prior belief. If the institute's released data shows that every woman with her characteristics has lung cancer, the guess becomes 100% (the posterior belief). The release has revealed her diagnosis.

Such an inference should fail the No Inference criterion, and the draft states the principle that makes it fail in paragraph 71: an inference counts only if it could not be obtained from general knowledge or from data about the population at large. It offers three ways to show whether an inference relies on a person's record. Ideally, the way in which the inference is drawn is traced (paragraph 71(a)). Where this is not possible, for example because the automated processing behind the inference is opaque, as is often the case for synthetic data, the inference is drawn from two versions of the given data, one produced with and one without the person's record, and the outcomes are compared (paragraph 71(b), footnote 43). We call this the leave-one-out comparison. Where neither is possible, a presumption

applies (paragraph 71(c)), to which we return below.

Neither of the first two routes reliably separates her case from population knowledge. Tracing shows which records an inference rests on, here the records of the women who share her characteristics, but not whether it reveals anything about her. Had these women all had high blood pressure, which is common among women of that age, tracing would give the same result, although the release would add far less to what could already be believed about her than in the case of lung cancer. Whether an inference discloses something depends on how far it departs from what holds for comparable people outside the data, and tracing does not ask this. The leave-one-out comparison fails for a different reason, which the draft's own examples show.

Example 12 describes a research institute holding five patient records in which date of birth and postcode have been generalised (Table 1). Example 19 aggregates the same records into a frequency table[1] of diagnoses by sex, year of birth and abbreviated postcode, and concludes that an individual's diagnosis can still be inferred by anyone who knows these three characteristics and that the individual's record was included. We agree. The three women born in 1951 form one cell of that table, and all three have lung cancer.

| Sex | Date of birth | Postcode | Disease |
|---|---|---|---|
| Male | **-**-1946 | 54*** | sclerosis |
| Male | **-**-1946 | 54*** | sclerosis |
| Female | **-**-1951 | 32*** | lung cancer |
| Female | **-**-1951 | 32*** | lung cancer |
| Female | **-**-1951 | 32*** | lung cancer |

**Table 1:** The given data of Example 12, reproduced from the draft Guidelines. The three women (shaded) form one cell of the frequency table in Example 19 and are used in Figure 1.

The leave-one-out comparison reaches the opposite result. Footnote 43 asks whether an inference about a person relies on that person's own record. It compares the released data, A(D), with the data the same anonymisation process would have produced without her record, A(D′). If the inference works in the same way for both, it is attributed to "the population at large". Applied to one of the three women, her record is removed and the cell still shows two women, both with lung cancer. The inference that she has lung cancer works exactly as before, so the leave-one-out comparison attributes it to the population at large, although it rests on two other patients. Footnote 44 then counts A(D′) as part of the prior belief, so the posterior cannot be "significantly different" from it (Figure 1).

This is the textbook case of attribute disclosure in statistical disclosure control, known as the homogeneity attack on $k$-anonymous data (Machanavajjhala et al., 2007). The size of the group does not settle the question on its own. Where the group is large and the same pattern holds among comparable people outside the released data, the inference is population knowledge and the draft's conclusion is right. Where it does not, the release has revealed the value, whether the group has three members or three hundred. The draft leaves this distinction to the reader: footnote 42 asks for a "large group" without indicating how large, and the leave-one-out comparison in footnote 43 gives the same answer whatever the size of the group. A comparison that removes one record at a time captures what can be learned from a person's participation in the data, not what can be inferred from other records that carry the same information (Kifer & Machanavajjhala, 2011). Applied to Examples 17 and 19, the leave-one-out comparison would reach the opposite conclusion to the draft's own, which treats exactly this situation as relating to each member of the group. In Example 17, a manager learns that all team members took part in a survey and none gave a score above 3, and the draft concludes that this information relates to each member individually. Paragraph 71(c), which allows a presumption where the original data has been deleted, relies on the same footnote 44, so a controller can again argue that the inference could have been drawn from the remaining members of the group (comment 5).

[1] The draft calls this aggregate a histogram. Since the diagnosis is a categorical variable, the aggregate is a frequency table (or, displayed graphically, a bar chart). See comment 6 in Section 3.

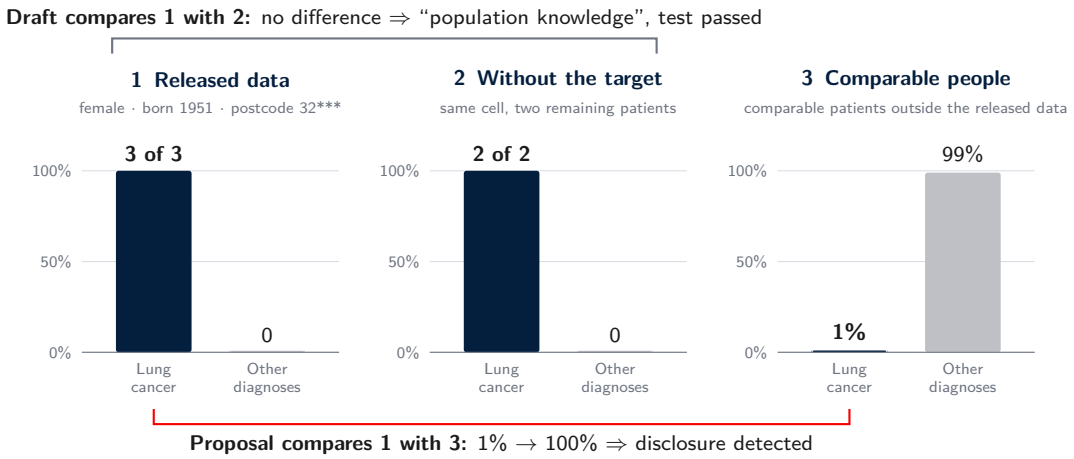



**Figure 1:** The three women in Table 1 all have lung cancer, and one of them is the target. The draft removes only her record (2), sees the same picture and treats her diagnosis as population knowledge. Compared with comparable people (3), the release raises the belief that she has lung cancer from about 1% to 100%. The 1% is illustrative.

In our applied work this is not a theoretical concern. When assessments are prepared against the draft, the population-knowledge argument is the most readily available route to concluding that the data is anonymous, and the draft gives the reviewer no basis to reject it. We propose that:

1. the prior belief be measured against comparable people, as defined above – for instance, how common lung cancer is among women born in 1951 in that region who are treated at comparable institutes – and not against the remaining members of the same group in the released data. A related idea underlies $t$-closeness in statistical disclosure control, which compares the distribution of a sensitive value within each group with its distribution in the data as a whole rather than with the group itself (Li et al., 2007). We propose a baseline outside the released data, because the data as a whole can share the pattern of the group;
2. where all or nearly all people in a group of the released data share the same sensitive value – as the three women who all have lung cancer do – the inference about each of them be presumed specific and meaningful unless the controller demonstrates at least one of the following:
   - the value is about as common among comparable people, citing a documented source for that figure; without such a source, this ground is not available;
   - no relevant entity can tell whether a given person belongs to the group, for example because the data is a sample whose members are not known (comment 9);
   - the released values have been altered enough that the pattern does not reliably reflect the original data.

Requiring that a release change nothing that can be believed about a person would rule out useful statistics altogether (Dwork & Roth, 2014). Measuring the prior belief against comparable people keeps the population knowledge that statistics are meant to provide, and it still detects the case in Figure 1. Detailed wording proposals are given in comments 3, 5 and 8 in Section 3.

# 3 Paragraph-by-paragraph comments

Comments are numbered for reference and ordered by paragraph number in the draft Guidelines, not by importance. Quoted text from the draft is set in italics.

| Original paragraph | Observation and rationale | Proposed paragraph |
|---|---|---|
| **1 · ¶ 41**<br>*"Equally, controllers should ensure adequate documentation of the anonymisation processing. This documentation of the anonymisation process (including for the testing of the supposedly anonymous datasets) makes it possible to demonstrate both the GDPR-compliance of the anonymisation process, and the effectiveness of the anonymisation itself."* | **Observation.** The paragraph does not say what adequate documentation of the testing contains, so assessments cannot be compared or re-run by a reviewer.<br>**Rationale.** A minimum content would make the effectiveness of anonymisation demonstrable in a reproducible way and allow supervisory authorities to review assessments consistently. | **Comment.** Indicate the minimum content of the documentation of testing. For each of the three criteria, the documentation should state the relevant entities and what they are assumed to know, the attributes treated as identifying and as sensitive, the inferences tested, the test performed with its measure and threshold, the result, and any residual risk accepted with its justification. |
| **2 · ¶ 57; Example 15**<br>*"the No Record Isolation criterion may involve looking at, for example, multiple lines in a table or database table"* | **Observation.** The draft gives one example of data collected over time (Example 15) but no specific guidance for such data. Under paragraph 56, every value in a record counts towards making it unique, so records with repeated measurements will often fail the No Record Isolation criterion. Paragraph 98 then asks whether a unique record can be matched with attributes the relevant entity knows. For data collected over time, an entity does not need to know the whole record: research shows that a few single observations within a record, such as a few places and times or a few diagnoses with dates, can be enough to single out a person (de Montjoye et al., 2013; Tamersoy et al., 2012), and AI agents can carry out such linkage effectively (Thees et al., 2026).<br>**Rationale.** Longitudinal data are common in health research, mobility and official statistics, yet there is no generally accepted method for assessing their disclosure risk, and this remains an active area of research. The Guidelines cannot settle this, but they can state what an assessment must at least address. | **Comment.** Clarify how paragraph 98 applies when a relevant entity knows only part of a record, and require the assessment to state and justify how many points, and which ones, an entity is assumed to know. |

| Original paragraph | Observation and rationale | Proposed paragraph |
|---|---|---|
| **3 · Footnote 42**<br>*"Here, 'information about the population at large' means any large group of people about whom information is not collected individually, but extrapolated from a sample, with any connection to the individuals left behind."* | **Observation.** This definition fits the prior belief we propose in Section 2.2: knowledge about a large group, extrapolated, with no connection to the individuals in the data. Footnote 43 does not apply it. A(D′) is produced from the same individually collected records, and where a group is homogeneous, the connection to its remaining members is not left behind. The definition also gives no indication of how large "large" is.<br>**Rationale.** This links the two footnotes and gives practitioners a baseline that cannot collapse into the remaining members of the same group. | **Comment.** Clarify that the comparison in footnote 43 must meet the definition in footnote 42: the prior belief should rest on information about a genuinely large group that was not collected from the individuals in the given data. If no numerical threshold is prescribed, identify the factors practitioners should consider. |
| **4 · Footnote 44**<br>*"Prior belief" (see footnote 40) ... "A(D′) in footnote 41" ... "significantly different posterior belief".* | **Observation.** Footnote 44 refers to "footnote 40" for the prior belief and to "A(D′) in footnote 41". The prior belief is introduced in footnote 42 and A(D′) is defined in footnote 43, so both references appear to be off by two.<br>**Rationale.** Readers cannot follow the comparison without the correct references. | **Comment.** Correct the references to footnotes 42 and 43. |
| **5 · Footnotes 43–44; ¶ 71(c)**<br>*Footnote 43: "if the inference on John Smith works in the same way for A(D) and A(D′), it is not an inference drawn from his data but from data relating to the population at large". Footnote 44: "Prior belief" "could include knowledge coming from an alternative version of the given data which has been produced without the record of the persons that are being identified".* | **Observation.** Where all members of a group in the released data share the same sensitive value, removing one of them leaves the pattern unchanged. The inference then passes the test in footnote 43 and, because footnote 44 counts A(D′) as prior belief, is treated as population knowledge, although it reveals the value of every group member. Example 17 is such a case. Paragraph 71(c) relies on the same footnote and sets no standard of evidence. See Section 2.2.<br>**Rationale.** The comparison in footnote 43 asks whether an individual's record was needed, not whether the release revealed their sensitive value. | **Comment.** Clarify that the prior belief must be measured against what could be known about comparable people without the released data, not against the remaining members of the same group. Where all or nearly all people in a group share the same sensitive value, presume the inference is meaningful unless the controller demonstrates at least one of the following: that the value is about as common in the comparable population, citing a documented source; that no relevant entity can tell whether a given person belongs to the group; or that the released values have been altered enough that the pattern does not reliably reflect the original data. Apply the same standard to the presumption in paragraph 71(c). |

| Original paragraph | Observation and rationale | Proposed paragraph |
|---|---|---|
| **6 · ¶ 81; Example 19**<br>*"The data from Example 12 is aggregated to create a histogram of diagnoses grouped by combination of sex, year of birth and abbreviated postcode."* | **Observation.** The diagnosis is a categorical variable. A histogram displays the distribution of a numeric variable over intervals; counts of a categorical variable by group form a frequency table, or a bar chart when displayed graphically.<br>**Rationale.** Correct terminology. Frequency tables are the object of established tabular data protection methods, which practitioners can then apply directly. | "The data from Example 12 is aggregated to create a frequency table of diagnoses grouped by combination of sex, year of birth and abbreviated postcode." |
| **7 · ¶ 82**<br>*"This would, in particular, be the case for AI models or synthetic data. Specific inferences can be made by querying (or, in the case of AI models, prompting) the given data with additional information to elicit new information about a particular individual."* | **Observation.** Paragraph 82 states that synthetic data and AI models can reveal information about individuals, but does not explain how to test for this. The test in footnote 43 compares the output with and without one person's data. For synthetic data, this would mean generating the data again without each person, one at a time, and repeating this several times, because every run gives slightly different results. For large datasets and complex generators this is demanding, and the effect of a single record is often hidden by the variation between runs, so the test can pass for lack of power.<br>**Rationale.** Without this, every controller will choose its own test, and assessments of synthetic data and AI models cannot be compared. | **Comment.** Indicate which practical tests are acceptable instead, for example simulated attacks that try to find out whether a person was in the original data or to guess their sensitive values, with particular attention to unusual individuals. One such test is RAPID (Risk of Attribute Prediction-Induced Disclosure; Templ et al., 2026): it assumes an attacker who knows some characteristics of a person in the original data, such as age, sex and postcode, uses the synthetic data to predict that person's sensitive value, and measures how much better these predictions are than guessing from how common each value is. Clarify that tests of whether a person was in the original data are not sufficient on their own: attribute inferences should be assessed against what could be known about comparable people without the released data (see comment 5). |

| Original paragraph | Observation and rationale | Proposed paragraph |
|---|---|---|
| **8 · ¶ 85, fourth indent**<br>*"The diversity of the data. This refers to the amount of differences between the data (e.g. if many individuals share the same attribute values, the data has a low diversity)".* | **Observation.** This factor applies to all three criteria, but the paragraph does not say which attributes it refers to or in which direction it works. Low diversity in the characteristics that could identify a person, such as sex, year of birth and postcode, protects them, because many people look alike. Low diversity in the sensitive information within such a group, for example when everyone has the same diagnosis, is exactly what allows an inference (see Section 2.2). Without this distinction, the paragraph could be read as suggesting that a homogeneous group is safe.<br>**Rationale.** This aligns the key factor with the established distinction between identity disclosure and attribute disclosure (see Section 2.1). | **Comment.** Distinguish the characteristics that could identify a person from the information that could be inferred about them. Clarify that low diversity in the former lowers the risk of record isolation and linkage, while low diversity in the latter, within a group of people who share the same identifying characteristics, makes a specific and meaningful inference more likely. |

| Original paragraph | Observation and rationale | Proposed paragraph |
|---|---|---|
| **9 · ¶ 94; Example 22**<br>*"... nobody who knows about the inclusion of an individual in the given data are among the relevant entities."* | **Observation.** In Example 22, the research institute concludes that no one who could get hold of the data knows which patients are in it, for example because friends and family are unlikely to receive the data after a leak. The draft says such conclusions must be "underpinned by careful reasoning", but does not say what evidence is needed. Where the data covers a whole group, such as all employees of a company or all customers of a shop, anyone who knows that a person belongs to the group also knows that the person is in the data. In sample surveys, entities usually do not know who is included because the sample members are not published. This protection fails where taking part is visible, for example through interviewer visits, or where the sample covers most of a small group.<br>**Rationale.** Whether someone knows that a person is in the data decides whether the inference in Examples 12 and 19 is possible. The conclusion that no one knows should therefore rest on evidence, not only on reasoning. | **Comment.** Specify what evidence is needed to conclude that no relevant entity knows who is included in the data, and name situations in which this should be presumed known, such as data covering a complete, closed or publicly known group (e.g. all employees of a company). This is also the evidence a controller needs to rebut the presumption in comment 5. |

# 4 References


Article 29 Data Protection Working Party. (2014). *Opinion 05/2014 on anonymisation techniques* (WP216, 0829/14/EN) (Adopted 10 April 2014). https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, *3*, Article 1376. https://doi.org/10.1038/srep01376

Duncan, G. T., & Lambert, D. (1986). Disclosure-limited data dissemination. *Journal of the American Statistical Association*, *81*(393), 10–18. https://doi.org/10.1080/01621459.1986.10478229

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, *9*(3–4), 211–487. https://doi.org/10.1561/0400000042

European Data Protection Board. (2026). *Guidelines 02/2026 on anonymisation* (Version 1). https://www.edpb.europa.eu/documents/guideline/guidelines-022026-on-anonymisation_en

Federal Committee on Statistical Methodology. (2005). *Report on statistical disclosure limitation methodology* (Statistical Policy Working Paper 22, 2nd version). Office of Management and Budget. https://apps.bea.gov/icsp/fcsm/assets/docs/SPWP22_rev.pdf

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Lenz, R., Naylor, J., Schulte Nordholt, E., Seri, G., de Wolf, P.-P., Tent, R., Młodak, A., Gussenbauer, J., & Wilak, K. (2026). *Handbook on statistical disclosure control* (2nd ed.). Center of Excellence on Statistical Disclosure Control. https://sdctools.github.io/HandbookSDC/Handbook-on-Statistical-Disclosure-Control.pdf

Kifer, D., & Machanavajjhala, A. (2011). No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data* (pp. 193–204). Association for Computing Machinery. https://doi.org/10.1145/1989323.1989345

Li, N., Li, T., & Venkatasubramanian, S. (2007). $t$-Closeness: Privacy beyond $k$-anonymity and $\ell$-diversity. In *Proceedings of the 23rd IEEE International Conference on Data Engineering (ICDE)* (pp. 106–115). IEEE. https://doi.org/10.1109/ICDE.2007.367856

Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). $\ell$-Diversity: Privacy beyond $k$-anonymity. *ACM Transactions on Knowledge Discovery from Data*, *1*(1), Article 3. https://doi.org/10.1145/1217299.1217302

Tamersoy, A., Loukides, G., Nergiz, M. E., Saygin, Y., & Malin, B. (2012). Anonymization of longitudinal electronic medical records. *IEEE Transactions on Information Technology in Biomedicine*, *16*(3), 413–423. https://doi.org/10.1109/TITB.2012.2185850

Templ, M. (2017). *Statistical disclosure control for microdata: Methods and applications in R*. Springer International Publishing. https://doi.org/10.1007/978-3-319-50272-4

Templ, M., Kowarik, A., & Meindl, B. (2015). Statistical disclosure control for micro-data using the R package sdcMicro. *Journal of Statistical Software*, *67*(4), 1–36. https://doi.org/10.18637/jss.v067.i04

Templ, M., Thees, O., & Müller, R. (2026). *RAPID: Risk of attribute prediction-induced disclosure in synthetic microdata* [Preprint, under review]. arXiv: 2602.09235.

Thees, O., Müller, R., & Templ, M. (2026). Agentic AI-powered re-identification: An emerging, scalable threat to mobility microdata privacy. In *Privacy in Statistical Databases* (pp. 319–332). Springer. https://doi.org/10.1007/978-3-032-37883-5_21