RESEARCH PAPER

Aggregation, Synthesisation and Anonymisation

A Call For A Risk-Based Assessment of Anonymisation Approaches

Published in Data Protection and Privacy: Data Protection and Artificial Intelligence, edited by Dara Hallinan, Ronald Leenes and Paul de Hert, CPDP Vol. 13, Hart Publishing 2021.

Sophie Stalla-Bourdillon*

Senior Privacy Counsel and Legal Engineer, Immuta

Professor of Information Technology Law and Data Governance,University of Southampton

Alfred Rossi**

Research Scientist, Immuta

Senior Lecturer, Ohio State University, Department of Computer Science and Engineering

IMMUTA™

This paper argues that the de-identification spectrum resulting from the common readings of both CCPA and the GDPR is oversimplified.

This is because in order to assess the output of a data transformation process, including aggregation, one should look beyond the output data and the technique applied on the input data: one should look at the data environment and the combination of both technical and organisational controls implemented to manage access to data.

We thus offer a new analysis of anonymisation controls and explain why this analysis is particularly useful in the context of data analytics and machine learning, where models can remember input data. This analysis applies even if decentralised techniques are available such as federated learning. Put simply, a similar approach can be applied to both what are traditionally thought of as a "dataset" and aggregate data products, such as summary statistics and models, which are key ingredients in producing synthetic data. What is more, we offer guidance for a more nuanced reading of both CCPA and the GDPR in order to effectively incentivise best data governance practices.

* Sophie Stalla-Bourdillon is Senior Privacy Counsel and Legal Engineer at Immuta and a professor of Information Technology Law and Data Governance at the University of Southampton.

** Alfred Rossi is a research scientist at Immuta, and a senior lecturer in the Ohio State University department of Computer Science and Engineering.

I. Introduction

2018 was the year of the General Data Protection Regulation (GDPR),[1] at least within the European Union (EU). While the GDPR was adopted in 2016 and EU Member States were required from this date to prepare the terrain for its application, its direct effect started on 25 May 2018.

From that date both public and private actors within EU Member States have been required to comply with the law.

2020 is the year when the California Consumer Privacy Act (CCPA) comes into effect.[2] Much has been written about both laws, and detailed comparisons[3] have already been released to support the work of compliance teams working for organisations operating in both regions. While the CCPA has certainly been influenced by the GDPR — in fact the language of some of its provisions is very close to the language found in the GDPR[4] — inevitably, differences have emerged. Whether these differences should be seen as merely language differences that should not prevent homogeneity of practices on the ground, or rather as conceptual differences that should lead to divergences of practices, is the big question.

This chapter deals with the way the material scope of this privacy or data protection legislation is defined and raises the question of what process is needed to transform personal information or personal data into non-personal information or non-personal data. We argue that anonymisation warrants a blended approach combining both context and data controls, even when the aggregation route is chosen. This should hold true although modern privacy or data protection legislations developed in different jurisdictions (ie, the US and EU) do not expressly refer to context controls. Context controls should thus be seen as implicit requirements.

To start with the Californian approach, the CCPA distinguishes between two categories of non-personal information, as per section 1798.140 of the California Civil Code[5]: de-identified information and aggregate information. The distinction seems to rely upon the assumption that aggregate information is always higher or further right on the de-identification spectrum[6] than de-identified information. Put simply, the distinction seems to rely upon the assumption that aggregate information is (much) safer than de-identified information, in terms of re-identification risks.

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/ 1–88.

[2] Note that Brazil's General Data Protection Law (GDPL) is also coming into force in August 2020, with Brazil's new data-protection agency to start working in October 2019. The GDPL is very similar to the GDPR both in substance and spirit.

[3] See, eg, OneTrust DataGuidance Staff and Future of Privacy Forum Staff, 'Comparing Privacy Laws: GDPR v. CCPA,' https://fpf.org/2019/12/18/comparing-privacy-laws-gdpr-v-ccpa/ (last accessed 4 March 2020).

[4] Consider for example the definition of pseudonymisation at Cal. Civ. Code §§ 1798.140(r), which borrows from the GDPR definition of pseudonymisation at Art 4(5) or the definition of business at Cal. Civ. Code §§ 1798.140(c)(1), which borrows from the GDPR definition of controller at Art 4(7).

[5] Cal. Civ. Code §§ 1798.140(a) and (h). For other US rules excluding aggregate information from the definition of personally identifiable information, see, eg, rules adopted by the Securities and Exchange Commission, 17 CFR §248.3 (u)(2)(ii)(B) ('Personally identifiable financial information does not include:.. Information that does not identify a consumer, such as aggregate information or blind data that does not contain personal identifiers such as account numbers, names, or addresses.')

[6] A scale between two extreme points (personal data and anonymised data) is usually used in the literature to explain the concepts of personal data, pseudonymised data, de-identified data, aggregate data and anonymised data. See, eg, Kelsey Finch, 'A Visual Guide to Practical Data De-Identification,' Future of Privacy Forum, https://fpf.org/2016/04/25/a-visual-guide-to-practical-data-de-identification/ (last accessed 4 March 2020).

A similar assumption could be said to underlie the GDPR, although the GDPR appears more restrictive than the CCPA.[7] While the GDPR does not expressly recognise the concept of de-identified data (it introduces the concept of pseudonymisation), the GDPR excludes from its remit anonymised data in fine. GDPR Recital 26 is usually used as the initial prong of an *a contrario* reasoning in order to derive the test for anonymised data.[8] In addition, GDPR Recital 162, which deals with certain types of processing activities, clearly specifies that '[t]he statistical purpose implies that the result of processing for statistical purposes is not personal data (including pseudonymised data), but aggregate data.'[9] It thus appears that the GDPR also draws a distinction between personal data and aggregates, albeit in its non-binding part.

In this chapter, we argue that the de-identification spectrum commonly used to explain why aggregates are not personal data is oversimplified. This is because in order to assess the output of a data transformation process, including aggregation, one should look beyond the output data and the technique applied on the input data: one should look at the data environment and the combination of both technical and organisational controls implemented to manage access to data. While re-identification scandals such as AOL's release of search terms or Netflix's sharing of movie recommendations because of poor de-identification methods have been heavily discussed,[10] aggregation failures have also been well-documented in the literature.[11]

After all, the US Bureau of the Census would not be looking to employ differential privacy if aggregation was sufficient.[12]

Building upon prior work,[13] we offer a new analysis of anonymisation controls and explain why this analysis is particularly useful in the context of data analytics and machine learning, where models can remember input data.[14] This analysis applies even if decentralised techniques are available such as federated learning.[15] Put simply, a uniform approach can be applied to both traditional datasets and aggregate data products, such as summary statistics and models, which are usually used to produce what is now called synthetic data.

[7] This is because recital 26 specifies that data that has undergone pseudonymisation should be deemed as personal data. ('Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person.')

[8] GDPR, Recital 26:
'To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.'

[9] GDPR, Recital 162.

[10] See, eg, Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) *57 UCLA Law Review* 1701; Article 29, Data Protection Working Party, Opinion 005/2014 on Anonymisation Techniques, adopted on 10 April 2014, WP216 (hereafter Art 29 WP Anonymisation Techniques.)

[11] In particular in the literature on differential privacy. See, eg, Cynthia Dwork and Aaron Roth.'The Algorithmic Foundations of Differential Privacy.' Foundations and Trends® in Theoretical Computer Science 9, no. 3–4 (2013): 211–407. https://doi.org/10.1561/0400000042.

[12] John M Abowd, 'Disclosure Avoidance and the 2018 Census Test: Release of the Source Code,' The United States Census Bureau, www.census.gov/newsroom/blogs/research-matters/2019/06/disclosure_avoidance.html (last accessed 4 March 2020).

[13] See, eg, Finch, 'A Visual Guide to Practical Data De-Identification' (n 13). See also Jules Polonetsky, Omer Tene, and Kelsey Finch, 'Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification' (2016) 56(3) Santa Clara Law Review 593.

[14] See, eg, Michael Veale, Reuben Binns, and Lilian Edwards, 'Algorithms That Remember: Model Inversion Attacks and Data Protection Law,' (2018) 376(2133) *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 20180083, https://doi.org/10.1098/rsta.2018.0083.

[15] Federated learning, sometimes called collaborative learning, trains data across multiple decentralised edge devices or servers. Training data is thus kept locally, never exchanged between devices or servers and are not holding local data samples, without exchanging their data samples. In other words, federated learning trains models against data that is separated in silos. The architecture for federated learning can vary widely. See, eg, Brendan McMahan et al, 'Communication-Efficient Learning of Deep Networks from Decentralized Data' [2017] *Artificial Intelligence and Statistics* 1273–82, http://proceedings.mlr.press/v54/mcmahan17a.html.

What is more, we suggest that a nuanced reading of both the CCPA and GDPR is preferable in order to effectively incentivise best data governance practices. While the definition of aggregate information under the CCPA does not expressly require a combination of technical and organisational controls, the regulatory goal is that at the end of the aggregation process aggregates are not linked or reasonably linkable to any consumer or household. Yet, this can only be achieved if, on top of the aggregation process itself, a combination of technical and/or organisational measures are implemented with a view to transform the data and control the data environment.

With regard to the GDPR, we suggest that the exclusion of aggregates from the remit of the regulation should not be systematic and at the very least should not be done on the basis of an irrebuttable presumption. Furthermore, data controllers, when producing aggregates, should assess the effectiveness of the combination of technical and organisational measures to properly characterise the output of the aggregation process and determine its legal effect.

Importantly, these suggestions should be followed to assess synthesisation processes and characterise their outputs. This is because, even if synthetic data is considered to be a valid alternative to original data,[16] at the end of the day synthetic data is data sampled from a model derived from aggregate data.

This chapter is organised as follows. In Section II, we unpack and refine the inference model that is commonly used to assess re-identification risks, taking into account four types of inference attacks in order to weigh the potential impact of anonymisation processes. The inference model is a general model of a knowledge-based attacker who is seeking to infer personal information that pertains to data subjects. This model is particularly useful for analysing the effectiveness of mitigating technical and organisational measures (ie, controls) in a machine learning context. It is on the basis of this model that we then analyse aggregation and synthesisation methods, highlighting their inherent limits in Section III. In Section IV we explain the concepts of data and context controls with specificity. In Section V, we compare anonymisation controls, and highlight the necessity of context controls. In Section VI, we draw the lessons from previous sections with a view to offer guidance for interpreting de-identification, aggregation or anonymisation provisions found in key privacy or data protection legislations such as the CCPA and GDPR. As the CCPA is representative of the US approach to de-identification as it builds upon the Federal Trade Commission's approach to de-identification,[17] and the US approach to anonymisation is usually opposed to the EU approach, considering both frameworks enable us to explore the potential for convergence.

[16] Neha Patki, Roy Wedge, and Kalyan Veeramachaneni, 'The Synthetic Data Vault,' in 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA) (2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, Canada: IEEE, 2016), 399–410, https://doi.org/10.1109/DSAA.2016.49.

[17] Staff, Ftc. 'Protecting Consumer Privacy in an Era of Rapid Change–A Proposed Framework for Businesses and Policymakers' [2011] *Journal of Privacy and Confidentiality*. https://doi.org/10.29012/jpc.v3i1.596; Simson L Garfinkel, 'De-Identification of Personal Information' (National Institute of Standards and Technology, October 2015), https://doi.org/10.6028/NIST.IR.8053.

II. How to Unpack the Inference Model

Our analysis makes use of the attack model used in differential privacy[18] (referred to herein as the *inference model*) as a general framework for analysing data privacy controls.

This construct is useful as it is sufficiently general to accommodate a generic knowledge-based attacker in relation to mitigation actions. Also, it can reveal latent assumptions which may lead to inaccurate or misleading comparisons of privacy techniques.

The medical data literature factors privacy (in the sense of confidentiality) risk into the product of two terms: *context risk* and *data risk*.[19] These terms arise naturally from viewing privacy attacks as probabilistic events which occur at some rate but may or may not be successful.

Viewing *attack* and *success* as separate but overlapping events, it follows from Bayes' theorem that the probability of a successful attack is the probability of success once an attack is occurring, scaled by the probability that an attack occurs. From an operational standpoint it is useful to think of attacks as events wherein a party accesses data for unauthorised purposes. A successful attack is then one under which an actor processing under an unauthorised purpose (ie, an attacker) possesses sufficient information to confidently infer the confidential information of a data subject.

Under this lens, controls focusing on limiting access and guarding against unapproved processing lower the odds of unauthorised processing, and therefore mitigate the attack event. Roughly speaking, attack event mitigations address the access *context*. Perhaps for this reason, the medical data literature refers to the corresponding risk as the *context risk*. In other words, the context risk refers to the likelihood of unauthorised access happening.

In addition, it is also possible to mitigate the success event (ie, relative to the probability that an attack is occurring, what is the likelihood that it is successful?). This risk is referred to in the medical data literature as the *data risk*. Since the data is already understood as being accessed, mitigations of this form focus on how to alter the data such that it proves only marginally valuable in enhancing the attacker's understanding of confidential information of individual data subjects.

It should be noted that we do not necessarily assume sophistication or intent on behalf of the attacker. An actor becomes an attacker when they process information under an unauthorised purpose. This includes unintentional processing for unapproved purposes, as well as inappropriate reliance on additional information, including unintentional reliance on prior knowledge resulting in inadvertent recognition. Further, an actor may not be a natural person and should be understood to include automated processes.

[18] See, eg, Dwork and Roth, 'The Algorithmic Foundations of Differential Privacy' (n 11) 6.

[19] See, eg, Khaled El Emam and Bradley Malin, 'Appendix B: Concepts and Methods for De-Identifying Clinical Trial Data, Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk' (National Academies Press (US), 2015), www.ncbi.nlm.nih.gov/books/NBK285994/; Khaled El Emam, Guide to the De-Identification of Personal Health Information, 1st edn (Auerbach Publications, 2013), https://doi.org/10.1201/b14764. See, for a generalization of this approach, Information and Privacy Commissioner of Ontario, 'De-identification Guidelines for Structured Data', (2016), www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf. See also ICO Code of Practice, Anonymisation: 'Managing Data Protection Risk Code of Practice', (2012), https://ico.org.uk/media/1061/anonymisation-code.pdf (last accessed 5 March 2020) [hereafter ICO, Anonymisation Code of Practice]; Elaine Mackey, Mark Elliot, Kieron O'Hara, The Anonymisation Decision-making Framework, (UKAN Publications, 2016), https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf.

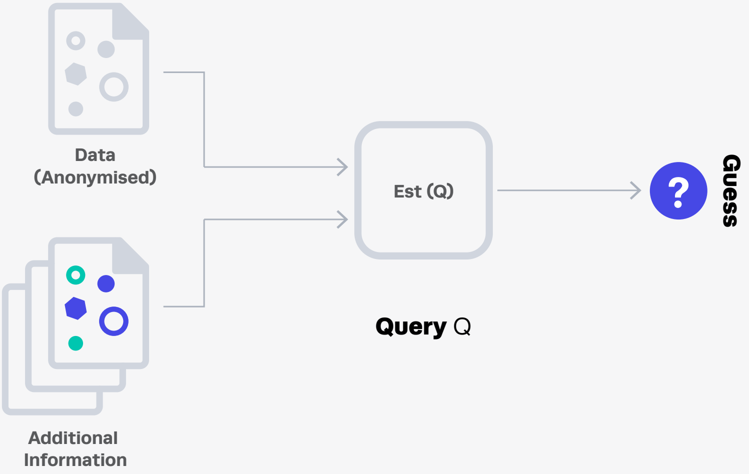

Diagram 1: A generic query setting aimed at protecting a 'dataset' style data.

The inference model (Diagram 1) formalises the thought process of an attacker who utilises a data product to answer a question. To this end, we think of the attacker as having a query, denoted Q. In answering the question, the attacker is allowed to formulate and execute a plan for guessing (estimating) the answer to Q. This plan, *denoted Est(Q)*, can incorporate both information obtained from the data product (via any modes of access available to the actor), as well as any *additional information* available to the actor.

For clarity, we now elaborate on the components of the model:

1. **The actor** is a process which aims to process data and formulate queries and their estimations. The actor is not depicted in Diagram 1.
2. **The Data** is the data product being accessed. It can be a dataset, query output from a model trained on private data, a white-box description of a model (eg, a neural network graph and the corresponding weights), etc. Access to data is constrained by the scenario and all access data is subject to any employed data transformation techniques. For instance, the actor may only be permitted authenticated access to data via database software that logs queries for auditing.
3. **The query** represents the analytic objective of the access, which may be to dump the data, train a model, compute some statistical aggregates, or even access a sensitive item from a patient's treatment history.
4. **The query estimate, Est(Q),** denotes the actor's plan for answering the query posed by Q. This plan (and its formulation) may make use of additional information available to the actor.
5. **A guess** comprises the output of the query estimate. A query may not be answerable with certainty; in this case the actor's plan is allowed to output a guess.
6. **Additional information** is simply information that the actor knows or has access to. It may, for example, include public or confidential information or data available to the actor, as well as the prior knowledge of the actor including the results of past queries.

The inference model naturally captures a number of common data access patterns. To name a few:

- access to static datasets;
- query access to databases;
- black-box (query) model access;[20]
- white-box model access; [21]
- differentially private access to data;
- interactive scenarios wherein

For the purposes of discussion, it is helpful to loosely categorise unwanted inferences along types.[22] In doing so we first outline the distinction between direct identifiers, indirectly identifying attributes, attributes, and tokens.

First, recall that an attribute is a piece of information associated with a record or an individual, which can either be unique to one record or common to several records.

Generally speaking, an identifying attribute is any attribute, knowable to an attacker, that can be associated with a natural person. An identifying attribute is a direct identifier when the attribute value is unique to an individual. All other identifying attributes are referred to as indirect identifiers. These are pieces of information (such as height, race, hair colour, etc) that can be used in combination to single out an individual's records.

A token (ie, the output of a masking process using a tokenisation method) replaces an identifying attribute or a non-identifying (but sensitive) attribute with a mathematically unrelated value through a transformation process that is difficult to reverse. A token replacing an identifier becomes identifying when the attacker possesses (or has access to) additional information that allows them to reverse the transformation.

Notably, even if the token is not reversed, a token that replaces an indirect identifier can still act as an indirectly identifying attribute if tokenisation is not performed by value (and not by record). In other words, a token can still act as an indirectly identifying attribute if the tokenisation method is homogeneously applied to the entire dataset.[23]

Crucially, over time it is possible that non-identifying attributes become indirectly identifying attributes due to the progressive enrichment of additional information.[24] Therefore, as explained below, simply removing direct and indirect identifiers will never be enough to mitigate once-and-for-all the four inferences mentioned above.

[20] The actor is not given a complete mathematical description of a model but may present it with test data for classification.

[21] The actor has access to a complete mathematical description of the model.

[22] Of note, Article 29 WP distinguishes between three types of re-identification risks: singling out, linkability and inference. While our terminology overlaps with Article 29 WP's terminology, we offer a more granular approach by distinguishing between participation inference and attribute inference. See Art 29 WP Anonymisation Techniques (n 10) 11–12.

[23] As an example, consider a consistent tokenisation of race, as this preserves the population statistics of the data, meaning an attacker could use demographic information to re-identify the underlying values.

[24] As explained by Steven M Bellovin et al, 'de-identification suffers from an aging problem'. Steven M Bellovin, Preetam K Dutta, and Nathan Reitinger, 'Privacy and Synthetic Datasets' (2019) 22(1) Stanford Technology Law Review.

We now describe the unwanted inference types, and shall sometimes refer to them as *inference attacks* in situations wherein such performances are not permitted:

1. **Identity inference:** Conclusion relating to the identity of an individual reached when considering direct identifiers, indirectly identifying attributes, and/or attribute values alone or in combination with additional information.

2. **Attribute inference:** Conclusion relating to the values of (sensitive) attributes attached to an individual record reached when considering direct identifiers, indirectly identifying attributes, and/or attribute values alone or in combination with additional information.

3. **Participation inference:** Conclusion relating to the participation of an individual to a data source reached when considering direct identifiers, indirectly identifying attributes, and/or attribute values alone or in combination with additional information.

4. **Relational inference:** Conclusion relating to the relationship or link between one or more individual records reached when considering direct identifiers, indirectly identifying attributes, and/or attribute values alone or in combination with additional information.

We now try to understand and mitigate the four kinds of inference attacks in the inference model.[25] Here, the goal becomes prevention of the production of query results that will lead to identity inference, attribute inference, participation inference or relational inference.

In order to assess the strength of the interference mitigation strategy, it is useful to conceptualise a query interaction that will aim at deriving information about the data. A typical query in this context could be 'whose record is it?' (identity disclosure), 'what is John's disease?' (attribute inference), or 'was the model trained on John's credit history?' (participation inference), or 'does this diagnosis relate to the same patient?' (relational inference).

In a generic setting where their actions are not known, an attacker should be prevented from using the data to enhance their ability to confidently make unauthorised inferences concerning the confidential information of data subjects. It should be noted that the attacker may *already* be able to accurately make such inferences based on additional information. It thus follows that if an attack is arbitrary, the focus must be on mitigating *the enhancement* (and not prevention) of the attacker's inferential abilities.

Formally, one may think of a generic guessing attack as a Bayesian process wherein prior knowledge is modelled by a probability distribution over guesses, with the probability of each guess reflecting the overall strength of the attacker's belief. The attacker's goal is then to consume query output in order to enrich their prior knowledge.

Note that the only known class of techniques that mitigate an attacker with access performing an arbitrary attack are differentially private.[26] These techniques are specifically designed to control the maximum amount of private information inferable from the data or query results information by an attacker, thereby hampering an actor's ability to significantly enrich their prior knowledge. This means that an attacker's guessing abilities

[25] See Diagram 1.

[26] Dwork and Roth, 'The Algorithmic Foundations of Differential Privacy' (n 11) 5. ('"Differential privacy" describes a promise, made by a data holder, or curator, to a data subject: "You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available."').

are only marginally enhanced in regard to the content or presence of any single record in the database. As such, their inferential ability remains essentially only as good as their prior knowledge and attackers wishing to learn something new are effectively prevented from doing so for *any* type of inference.

Useful applications of differential privacy work by introducing randomisation into an analysis.[27] The randomisation obscures the contribution of any single record by ensuring that analysis results are essentially equally likely over any pair of databases that differ by the presence of a single record. Such protections come at a cost, however, as any useful analysis must now be robust to noise. The intentional use of randomisation may seem prohibitive, but it turns out that large classes of problems in machine learning remain efficiently learnable, though typically with some reduction in accuracy that can be overcome with additional data.

A crucial secondary goal, then, is to design mitigation measures which maintain utility (ie, allow for good query estimation when the queries pertain to authorised purposes) yet ensure that any misuse of the data access, say by attempting to reverse the anonymisation, gives inconclusive results. This is not hopeless: in terms of the preceding example, a good differentially private analysis gives results with bounded error, yet it follows from the definition of differential privacy that access to the results only marginally increases the attacker's confidence in their attempts to guess confidential information.

The analysis of non-differentially private techniques typically requires a fixed attack scenario. In these analyses, the set of possible attacks executed by an attacker remains limited. The goal becomes to design mitigating measures that mitigate the odds of a successful attack when the full scope of controls is taken into account. Controls are essentially a means to mitigate risks. As we will explain in Section IV, they can be either data or context. Competing sets of controls are then meaningfully and directly comparable in terms of their respective probabilities of success.

For fixed attack scenarios, the question thus becomes, 'given a query, how much can one obtain identity, attribute, participation or relational inferences based on their access to query results information?' Using a scale from 0 to 1, it is possible to produce a quantified estimate of the inference risk taking into account all four types of inferences by probability of success.

Note that comparing controls based on the attack success probability does not work for differentially private techniques, since this requires consideration of a specific attack. After all, the attack may be based on extensive prior knowledge. This does not mean that differentially private techniques are inferior to other techniques. In fact, quite the opposite, as they guarantee that an attacker's ability to guess only marginally improves upon additional information. In other words, the attacker is guaranteed to learn little to nothing, and it is preferable to rank differentially private techniques in terms of the maximum amount of per-individual information inferable from the output.

That said, risk-based evaluation remains possible under scenarios where the set of possible attacks is fixed. Roughly speaking, any attack with low probability of success is guaranteed to remain low despite the possible incorporation of the differentially private results information. This works because the amount of usable information present in a differentially private release is bounded, and thus so must be the accompanying reduction in uncertainty of a guessing adversary.

The inference model, as described in Section II, is particularly useful to explain the limits of aggregation and synthesisation methods and compare effective mitigating technical and organisational measures (ie, controls).

[27] It can be shown that deterministic differentially private analyses, ie ones that do not employ randomisation, must always give the same output, and are therefore not useful.

III. How to Conceptualise Aggregation and Synthesisation

A. Traditional aggregation

Aggregated data is summary data. Simply put, aggregated data is metadata that serves to summarise data. Aggregated data is typically produced for statistical purposes (eg, to measure characteristics of population), but the term can be used here expansively to encompass other purposes such as processing, compression (eg, to reduce the size of storage), visualisation, or producing a model. In other words, aggregates need not be statistical in nature.

Aggregation operations are common for dataset data. For example, computing maximum (or minimum) value for a particular column within a dataset (eg, what is the *maximum (minimum)* salary for the population contained in the data set?), the *average* value for a particular column within a dataset (eg, what is the average salary for the population contained in the dataset?), or the *count* of records with a specified value for a particular column (eg, how many individuals have a salary of X within the population contained in the dataset?) are all processes that will produce aggregated data. The process of producing aggregated data is referred to as data aggregation or, often, aggregation. It should be distinguished from the process of de-identification, which is usually understood as the processing of stripping identifiers away, both direct identifiers and indirect identifiers, while keeping the data at the individual or event level.

Aggregation should also be distinguished from the process of generalisation, which consists in replacing a value with a less specific but semantically consistent value (eg, replacing a specific age with an age range within a record).

At present, there appears to be an assumption in privacy and data protection regulations that aggregate information is safe, as efforts to organise anonymisation techniques consistently list aggregation higher (or further right) on the de-identification spectrum than de-identified information. This assumption is known to be incorrect, and solving the problem of rendering aggregate data safe partially motivated the development of differential privacy.[28] For aggregate output to be deemed safe it is necessary to prevent an unauthorised third-party from being able to learn (infer with high confidence) personal information as regards data subjects among the aggregation input.

Mathematically, aggregate functions are functions that output a set of numbers derived from a set of inputs. At first glance this may indeed seem safe: after all, in reducing a database to a single number, a lot of information is thrown away. However, as we shall see, not only is this not necessarily the case, this is also not a sufficient condition to ensure that the data cannot be attributed to an individual.

[28] See Cynthia Dwork et al, 'Calibrating Noise to Sensitivity in Private Data Analysis,' in Shai Halevi and Tal Rabin (eds) *Theory of Cryptography*, vol 3876 (Springer Berlin Heidelberg, 2006), 265–84, https://doi.org/10.1007/11681878_14; Cynthia Dwork et al, 'Calibrating Noise to Sensitivity in Private Data Analysis,' (2017) 7(3) Journal of Privacy and Confidentiality 17–51, https://doi.org/10.29012/jpc.v7i3.405.

For example, the *maximum* aggregate must discard a lot of data, as its output only critically depends on rows achieving the maximum value, with all other values in the database remaining irrelevant. Yet when evaluated over company salary data, there may likely be only one row achieving the maximum corresponding to the company chief executive. It follows that an attacker with access to the maximum aggregate may have a good guess as to the executive's salary. One may think that the problem is due to the fact that this aggregate must depend upon more than a single data subject's information. Consider instead an aggregate that *encodes the entire database into a single number* using Gödel numbering.[28] The aggregate result depends on every record, yet it remains possible to infer the exact contents of any record through repeated division.

As we now outline, it is not necessary to resort to such extreme examples. Consider the following plausible participation inference attack demonstrating that the protections afforded by the *average* aggregate do not guarantee privacy.

Suppose that it is suspected that a local surgeon has a certain rare health condition. It is known that this disease occurs uniformly at random across the population, though a medical study is performed to see if there is a relationship between better outcomes and various socioeconomic attributes. The study publishes its findings, which are irrelevant for our purposes, but reports that all 100 positive local individuals participated, and, the average participant had an income of $58,720, and that one individual making more than $100,000 per year participated. It is well known from census data that individuals in this region have an average salary of $52,175 with a standard deviation of +/– $5,385, with only surgeons making more than $250,000 per year. By Chebyshev's inequality, the odds of observing an average salary that is more than one standard deviation higher due to chance alone is less than 1 per cent for the study size of 100. This suggests to the attacker that the observed shift in the mean is not due to chance. It is likely that the participating outlier is an extreme outlier. Moreover, it gives the attacker an estimate of the outlier's income: as 100*($58,720) – 99*($52,175) = $706,675, and the attacker can even further work out error bounds.

One interesting observation from the example above is that even though the average aggregate, unlike the maximum aggregate, depends on every value, it also tends to favour the privacy of certain individuals over others by responding disproportionately to outlying records.

Indeed, the problem seems to be related to how much an aggregate is influenced by the addition (or removal) of individual records. A goal, then, is to make aggregates safe by rendering them insensitive to the presence or absence of any individual input row. However, if made perfectly safe, the aggregate would be entirely insensitive to its input, and therefore fail to summarise it. This is problematic as safety is at odds with utility, and unconditional safety is, at best, aspirational.

This situation is easily remedied through differential privacy, which serves to guarantee that the resulting data product (eg, a model or query results information) does not carry more than epsilon bits[30] of personal information from any record. This ensures that the data product is of marginal value for enriching inference attacks provided that epsilon remains small.

[29] Eg, through reinterpreting the underlying bits of a record as a number, and then employing Gödel numbering to return a single number whose factorisation reveals the bits of any record. See Kurt Gödel, 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I,' Monatshefte für Mathematik und Physik 38–38, no. 1 (December 1931): 173–98, https://doi.org/10.1007/BF01700692.

[30] It should be noted that epsilon may be fractional, for instance epsilon = 0.001, which corresponds to situations wherein one would expect to need about 1,000.

However, a concern remains that each differentially private response on the same database may leak different information about the same individuals. Situations wherein vast repositories of differentially private results can be referenced or generated should be avoided. Of particular concern are adaptive attackers who may issue many differentially private queries with the goal of trying to collect as much information as possible about a group of targets. Put differently, *release context matters*. Controls must be put in place to prevent an adversary from using many aggregate summaries in conjunction to significantly enrich their knowledge of specific participants.

It should be noted that controls are not necessarily mutually exclusive with open data. After all, requiring authenticated access to data and making access to data contingent on access agreements only strengthens protection.

B. Synthesisation

Synthetic data is data drawn from a model which has been trained on real data. The model is generative in the sense it outputs (generates) something it believes to be consistent with the training data. Despite the appeal of using data to which no natural person corresponds, several problems exist. The generative model is trained on real data and thus is derived from the private, and perhaps sensitive, information of individuals. The question then is to what extent an attacker may be able to make inferences about the participation (or attributes) of these individuals from the behaviour of the model. In the limit of a large number of samples, it is often possible to reconstruct the parameters of the generative model with high fidelity, yielding several precise estimates of aggregate quantities derived from private data.

Synthetic data, without differential privacy, can be a very weak option. To see why, consider the *local surgeon* example of the previous section. A faithful synthetic data model will produce values for the income attribute which statistically agree with the income distribution as seen in the study data: the mean and the standard deviation. Even model parameters are not given directly to the recipient, with enough synthetic data it is possible to estimate both the mean and standard deviation to any desired precision. Given a high-quality estimate of the mean, the attack given in Section III A can be carried out. Namely, an attacker with knowledge of the size of the study (which is public) and an estimate of the mean income can infer the participation of the local surgeon in the training data, and thereby now possess compelling evidence that the local surgeon has a disease, as well as an estimate of the surgeon’s income for the year of the study. Note that this is possible *even though all data comes from a synthetic model*.

To mitigate such attacks, it is important to employ controls to limit the amount of personal information that is inferable from such quantities. Again, the natural family of techniques come from the field of differential privacy where such methodology is guaranteed to limit the number of bits of personal information that flow into such aggregates, making aggregate quantities less useful for making inferences.

What Section III shows is that both aggregation and synthesisation are not effective controls per se, despite the belief widely shared within the compliance community that both should automatically put the data outside the scope of privacy and data protection laws.

IV. How to Assess Aggregation Outputs

As hinted above, it is not enough to consider the output of the aggregation process to conclude that the re-identification risk is remote[31] or very small[32] and thereby declare that data usage should not be restricted anymore.

Considering the process through which the aggregated output has been obtained is essential, as aggregation processes vary in terms of privacy protections. As outlined in Section II, this should lead us to distinguish two types of controls, *data controls* and *context controls*, which affect either the data itself or its environment.[33]

As aforementioned, controls are essentially a means to mitigate risks. Different types of means can be used: technical means (such as data transformation techniques) or organisational means (such as contracts imposing obligations on data owners and data users, or policies specifying business processes within organisations acting as data owners or data users).

As explained below, in order to be truly effective, controls have to be combined. By effective controls we mean controls that lower the overall risk, both in terms of context and data risk, under at least one attack scenario as explained in Section II. What is more, aggregation should not be seen as an effective control. Mitigation of re-identification risks only happens after several steps are taken and aggregation is only one step in this process. Aggregation will in fact have to involve the implementation of both data and context controls to effectively mitigate re-identification risks.

Data controls are technical measures aimed to strengthen the protection of the confidentiality of the information. The strongest data controls are those that offer data stewards formal mathematical guarantees so that they are able to say to individuals: 'You will not be materially affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available.'[34] Differentially private methods, as explained below, are therefore the most obvious type of data control data stewards should be thinking about when wanting to produce aggregates.

Notably, differentially private aggregation can be undertaken through two routes, which are not necessarily equivalent in terms of degree of protection: global differential privacy and local differential privacy, as illustrated in Table 1. It should be noted that when using machine learning techniques to create models, both routes are worth exploring and likely to require fine-tuning over time.[35]

[31] ICO Code of Practice, 6 ('The DPA does not require anonymisation to be completely risk free — you must be able to mitigate the risk of identification until it is remote'). Ex-Article 29 WP writes that the 'the "means … reasonably to be used" test is suggested by the Directive as a criterion to be applied in order to assess whether the anonymisation process is sufficiently robust, i.e. whether identification has become "reasonably" impossible.' Art 29 WP Anonymisation Techniques (n 10) 8. As mentioned, CCPA defines aggregate consumer information as information that is 'is not linked or reasonably linkable to any consumer or household.' Cal. Civ. Code § 1798.140(a).

[32] This is the legal standard found in the US Health Insurance Portability and Accountability Act of 1996 (HIPAA). See 45 CFR § 164.514(b)(1).

[33] For a conceptualisation of the data environment see, eg, Mackey, Elliot, and O'Hara, *The Anonymisation Decision-making Framework* (n 19)

[34] Dwork and Roth, 'The Algorithmic Foundations of Differential Privacy' (n 11) 5.

[35] While it is known that large classes of efficiently learnable problems in machine learning remain efficiently learnable under differential privacy, significant barriers exist in the adoption of globally differentially private methods due to their reliance on specialised methods.

As it has been suggested:

A good technique for preventing model inversion attacks is simply keeping unnecessary data out of the training set. First, the data scientist should build a version of the model without differential privacy. (She should not release the model to the public at this stage.) She would note its baseline performance and then throw away the model. She would then iteratively build models with more noise until she reaches a minimum acceptable threshold for performance, or a maximum acceptable threshold for privacy loss. Assuming, then, that the privacy loss is acceptable, she could release the model into production.[36]

DATA CONTROLS	**DESCRIPTION**
Global differential privacy (GDP)	GDP is a technique employing randomisation in the computation of aggregate statistics. GDP offers a mathematical guarantee against identity, attribute, participation, and relational inferences and is achieved for any desired 'privacy loss.'[37]
Local differential privacy (LDP)	LDP is a data randomisation method that randomises sensitive values. LDP offers a mathematical guarantee against attribute inference and is achieved for any desired 'privacy loss.'[38]

Table 1: Examples of data controls for producing aggregated data

While some commentators have argued that differential privacy methods do not leave any room for utility,[39] the trade-off between utility and confidentiality is in fact context dependent. Assuming that the data is well-sampled and that there are no outliers, a satisfactory degree of utility should be reached. By way of example, any problem that is learnable in the probably approximately correct model (or PAC learnable),[40] remains PAC learnable under differential privacy.[41] Further, technologies such as TensorFlow Privacy augment machine learning methods with global differential privacy.[42]

[36] Sophie Stalla-Bourdillon, Alfred Rossi, and Gabriela Zanfir-Fortuna, 'Data Protection by Process: How to Operationalize Data Protection by Design' (2019), https://fpf.org/2019/12/19/new-white-paper-provides-guidance-on-embedding-data-protection-principles-in-machine-learning/.

[37] Matthew Green, 'What Is Differential Privacy?' A Few Thoughts on Cryptographic Engineering (blog), June 15, 2016, https://blog.cryptographyengineering.com/2016/06/15/what-is-differential-privacy/ (last accessed 5 March 2020).

[38] See, eg, Stanley L Warner, 'Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias' (1965) 60(309) *Journal of the American Statistical Association* 63–69, https://doi.org/10.1080/01621459.1965.10480775. See also Shiva Prasad Kasiviswanathan et al, 'What Can We Learn Privately?' (2011) 40(3) SIAM Journal on Computing 793–826, https://doi.org/10.1137/090756090.

[39] See, eg, Matthew Fredrikson et al, 'Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing,' in Proceedings of the 23rd USENIX Conference on Security Symposium, SEC'14 (San Diego, CA: USENIX Association, 2014), 17–32. (finding utility and privacy mutually exclusive in regard to warfarin dosing studies); id. at 29 ('[F]or values that protect genomic privacy, which is the central privacy concern in our application, the risk of negative patient outcomes increases beyond acceptable levels.')

[40] LG Valiant, 'A Theory of the Learnable,' Communications of the ACM 27, no. 11 (November 5, 1984): 1134–42, https://doi.org/10.1145/1968.1972.

[41] Kasiviswanathan et al, 'What Can We Learn Privately?' (n 38).

[42] See H Brendan McMahan et al, 'A General Approach to Adding Differential Privacy to Iterative Training Procedures,' ArXiv:1812.06210 [Cs, Stat], 4 March 2019, http://arxiv.org/abs/1812.06210.

Context controls are technical or organisational measures implemented to strengthen the protection of the confidentiality of the information queried, with no direct impact upon the content of the query results. Instead, external controls can have a direct impact upon who is able to formulate a query (eg, role- or attribute-based access control, data sharing agreement), how many queries a data user will be able to formulate (eg, query monitoring), the purposes for which the query results can be used (eg, purpose-based access control, data sharing agreement), and the mitigation actions the data curator and the data user will have to perform when aware that the re-identification risk is increasing given changes in the data environment. Context controls are illustrated in Table 2.

CONTEXT CONTROLS	TYPOLOGY	DESCRIPTION
Access control (RBAC, ABAC)	Technical control	Access rights are granted to data users through either the allocation of roles (function within organisation, department, team) or/and the use of policies which combine attributes. Policies can use any type of attributes (user attributes, data source attributes, column attributes, etc).
Purpose-based access control	Combination of technical and organisational controls	Purpose-based access control forces the data user to acknowledge the purpose under which she is requesting access to the data and requires the data user to agree with accessing the data for this purpose only. The purpose for which data is to be accessed can be expressly mentioned within the data sharing agreement concluded between the data curator and the data user or within an internal policy if the data curator and the data user belong to the same organisation.
Prohibition of linking	Combination of technical and organisational controls	In order to reduce the likelihood of all types of inferences, data users are prevented from linking data sources together (through technical means and data sharing agreement and/or internal policy).
Prohibition of re-identification	Organisational control	In order to prevent re-identification from happening, data users can be subject to an obligation not to re-identify individuals to which the information pertains (eg, through a data sharing agreement or an internal policy).
Prohibition of data sharing	Organisational control	The data user (ie, the recipient of the data) is under an obligation not to share the data with other parties.
Query monitoring	Technical or combination of organisational and technical controls	Query monitoring is facilitated by the query interface and is performed in real time by a compliance personnel or auditor. Query monitoring can also be automated through a privacy budget.

CONTEXT CONTROLS	TYPOLOGY	DESCRIPTION
Query termination	Technical control or combination of technical and organisational controls	Query termination is facilitated by the query interface and is performed in real time by a compliance personnel or auditor. Query termination can also be enforced automatically when the data user exhausts the privacy budget.
Obligation to monitor additional information	Organisational control	The curator is under an obligation to monitor publicly available information to assess the strength of the anonymisation process.
Obligation to comply with breach mitigation plan	Combination of technical and organisational controls	Each stakeholder (ie, the data curator and the data user) is under an obligation to take immediate mitigation action if they are aware of significant changes within the data environment (eg, the data curator terminates access to the data or the data user reports to the data curator) and could also be under an obligation not to contact re-identified (or likely to be re-identified) individuals. These duties are usually formulated within a data sharing agreement or within an internal policy.

Table 2: Examples of context controls for producing aggregated data

What Section IV suggests is that there exist a variety of controls that are relevant for lowering re-identification risks, be it through both the aggregation or the de-identification route.

This explains why an output-based approach to characterise aggregate data is not enough: a process-based approach is key, which should start by assessing the variety of data and context controls applicable or applied to the data and its environment.

The same holds true for de-identification: only when a process-based approach is adopted should it be possible to characterise the output data.

And the production of synthetic data is no exception to this consideration. After all, synthetic data is just data sampled from a model derived from aggregate data and with enough samples it is possible to reconstruct the model parameters and, therefore, learn everything that is inferable from the aggregate value, possibly including confidential information. This thus leads us to compare anonymisation controls in Section V and suggest that context controls are a must-have.

V. How to Compare Anonymisation Controls

Prior attempts to create representations of de-identification and/or anonymisation solutions have relied upon a two-dimensional spectrum or staircase.[43] Garfinkel, for example, explains that:

> all data exist on an identifiability spectrum. At one end (the left) are data that are not related to individuals (for example, historical weather records) and therefore pose no privacy risk. At the other end (the right) are data that are linked directly to specific individuals. Between these two endpoints are data that can be linked with effort, that can only be linked to groups of people, and that are based on individuals but cannot be linked back. In general, de-identification approaches are designed to push data to the left while retaining some desired utility, lowering the risk of distributing de-identified data to a broader population or the general public.[44]

While this presentation makes sense at a high level, it does not directly enable decision-makers to actually choose among data and context controls.

Omer Tene, Kelsey Finch and Jules Polonetsky go one step further in their attempt to provide effective guidance to both data users and compliance personnel and bridge the gap between technical and legal definitions.[45] They distinguish between pseudonymised, protected pseudonymised, de-identified, protected de-identified and anonymous data and introduce the concept of non-technical safeguards and controls, which are added to data modification techniques to produce either protected pseudonymised or protected de-identified data:

> Non-technical safeguards and controls include two broad categories: 1) internal administrative and physical controls (internal controls); and 2) external contractual and legal protections (external controls). Internal controls encompass security policies, access limits, employee training, data segregation guidelines, and data deletion practices that aim to stop confidential information from being exploited or leaked to the public. External controls involve contractual terms that restrict how partners use and share information, and the corresponding remedies and auditing rights to ensure compliance.[46]

In that model, aggregated data are considered to be safer than de-identified data and are not described as requiring non-technical safeguards and controls. While this spectrum should certainly be welcome in that it makes it clear that both technical and non-technical safeguards and controls are relevant for assessing the

[43] See, eg, Garfinkel, 'De-Identification of Personal Information' (n 17); Polonetsky, Tene, and Finch, 'Shades of Gray' (n 13).

[44] Garfinkel, 'De-Identification of Personal Information' (n 17) 5.

[45] Polonetsky, Tene, and Finch, 'Shades of Gray' (n 13).

[46] ibid, 606.

robustness of data modification processes, it relies upon a simplified binary conception of aggregated data for which non-technical safeguards and controls as well as other types of context controls do not seem to be needed. As explained in the introduction, though, such an approach is not surprising, as it seems to underlie several pieces of legislation.

What is more, this spectrum does not specifically locate synthetic data, which is increasingly seen as a valid alternative to aggregation.

Finally, the spectrum is not particularly adapted to an interactive query setting, which should make it possible to compare the robustness of data controls, with a view to select effective data and context controls for the use case at hand.

What Runshan Hu et al show is that, assuming sanitisation techniques are to be combined with contextual controls, in order to effectively mitigate the three re-identification risks identified by ex-Article 29 Data Protection Working Party, a two-dimensional representation of sanitisation techniques through the means of a spectrum is not necessarily helpful, as a different mix of sanitisation techniques and contextual controls could in fact be seen as comparable.[47] While they note that the only sanitisation technique able to mitigate on its own the three re-identification risks studied (ie, singling out, inference, linkability) is differential privacy on the condition that at least additional security controls are implemented, they do not assess the effect of synthesisation.

Bellovin et al only focus upon synthetic data and make it clear that a distinction should be drawn between vanilla synthetic data and differentially private synthetic data. However, they do not offer a comparative diagram for the different types of data controls and do not consider context controls.[48]

The primary difficulty in developing a coherent ranking of privacy techniques is that their effectiveness is highly context dependent. For instance, a medical study may involve exchanging detailed patient data with a partnering hospital in order to develop new treatments or identify subjects for clinical trials. The privacy concerns and controls employed for medical data will undoubtedly differ from those that collect browsing history to build and sell models for marketing purposes. And yet, despite these differences, the toolset in either case is essentially the same. Data risk mitigations include the familiar techniques of tokenisation, k-anonymisation, and local and global differential privacy. Context risk mitigations still include access controls, contracts, training, and auditing.

In both of the preceding examples it remains necessary to ensure that the privacy risk is tolerable. Despite the shared goal and common set of tools, implementations look very different. In the medical data example, it is assumed necessary that sensitive information be exchanged in order to fulfil the purpose. This means that it is not possible to prevent the receiving organisation from making sensitive inferences about the data subjects, after all this is the point. Thus, there is a heavier reliance on more costly context controls, including access and control restrictions, data sharing agreements, training for employees, etc.

[47] Runshan Hu et al, 'Bridging Policy, Regulation and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR,' in *Data Protection and Privacy: The Age of Intelligent Machines* edited by Ronald Leenes Rosamunde van Brakel, Serge Gutwirth and Paul De Hert. (Oxford: Hart Publishing, 2017), 115-142.

[48] Bellovin, Dutta, and Reitinger, 'Privacy and Synthetic Datasets' (n 24) 37, 41. ('For synthetic data, this means that without adding privacy-preserving features like differential privacy, there still remains risk of data leakage.')

The success of knowledge-based attacks is generally limited by two things: the availability of information and processing abilities of the attacker. It follows from this that anonymisation is context dependent. For instance, consider an attacker with intimate medical knowledge of a specific individual. Such knowledge, to the extent that it is not apparent from observation of the individual or their behaviour, would not be considered identifying and would be left in the clear in some approaches to anonymisation. Therefore, such information would not be anonymised to an attacker with access to the subject's medical records.

It is difficult to produce a comparative classification of anonymisation controls without consideration of situational specifics. Attempts to do so fail to acknowledge the role of both technical and non-technical controls and ultimately rely upon a gross oversimplification of the attack scenario that fails to adequately characterise the behaviour of the attacker. Moreover, ignoring situational differences leads to invalid comparisons of the guarantees of formal attack models that may not equally apply.

Let us go back to our generic query setting illustrated by Diagram 1 where we want to protect, through de-identification or aggregation, data to prevent query results from leaking information leading to identity inference, attribute inference, participation inference, or relational inference.

Let us further assume that all direct identifiers are masked in a way that is irreversible to the attacker. This can be achieved through nulling, hashing with salts, or encrypting direct identifiers with state-of-the-art techniques. In other words, let us assume that the first data control applied on the data is pseudonymisation. Pseudonymisation is often the first step towards both de-identification and aggregation and does not as such mitigate against any type of inference, as it leaves indirect identifiers or indirectly identifying attributes as they are, in the clear.[49]

Let us now consider the following five data controls:

Control 1: A system limits the allowed queries to aggregate queries and returns GDP-protected results.

Control 2: A system uses LDP on sensitive non-identifying attributes.

Control 3: A system uses k-anonymisation[50] on the indirect identifiers (but not on non-identifying attributes).

Control 4: A system nulls/tokenises, through hashing with salts[51] or encryption[52] (per record), indirect identifiers.

Control 5: A system aggregates attribute values.

These five data controls can thus be organised according to their robustness (ie, their formal or mathematical resilience towards the four types of inferences aforementioned).

[49] See Art 29 WP Anonymisation Techniques (n 10) 20–21.

[50] Bellovin, Dutta, and Reitinger, 'Privacy and Synthetic Datasets' (n 24) 37, 41. ('For synthetic data, this means that without adding privacy-preserving features like differential privacy, there still remains risk of data leakage.')

[51] See, eg, Pierangela Samarati and Latanya Sweeney, 'Protecting Privacy When Disclosing Information: K-Anonymity and Its Enforcement through Generalization and Suppression,' 1998, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.37.5829.

[52] Hashing with salts will usually be preferred to nulling in scenarios in which it is possible to effectively segment data users and it is a requirement that at least one segment should be able to link records together.

Control 1: With a suitable epsilon and limited number of queries, this data control mitigates all four types of inferences. In other words, when control 1 is combined with two context controls (ie, query monitoring and query termination, which can take the form of an automated privacy budget or ad hoc monitoring and query termination), it is superior to controls 2–5.

Control 2: With a suitable epsilon and limited number of queries, this data control releases data that is safe from attribute inference on sensitive attributes.

Control 3: With a suitable k, this data control releases data that is safe from identity and relational inference at t=n, but not future proof against future attacks if additional information evolves over time and is enriched. Query monitoring and query termination are not relevant to make control 3 offer a formal guarantee against attribute or participation inference. However, control 3 could then be combined with other techniques, such as control 2, to mitigate against attribute disclosure, for example. The advantage of control 2 over other techniques such as l-diversity[53] and t-closeness[54] is that control 2 guarantees that only a limited amount of private data is transferred.

Control 4: This control only mitigates identity and relational inference at t=n, but it is not future proof against future attacks if additional information evolves over time and is enriched, and in any case, it does not mitigate against attribute and participation inference. It is thus weaker than GDP, which is future proof against all forms of inferences.

Control 5: This control does not mitigate against any type of inference without additional controls, such as controls 1–4. Crucially, this is also true for synthetic data. What is obvious, though, is that in several instances aggregating sensitive attributes is a better option than keeping sensitive attributes in the clear.

From this description it becomes clear that both aggregation and synthetic data per se do not offer any means to mitigate against the four types of inferences aforementioned.

What Section V shows is that some forms of context controls are always needed. Ultimately the choice of the data control will depend upon the utility requirements of the use case, but inevitably the weaker the data controls, the stronger and the more diverse the context controls will have to be. The drawback of relying upon a great variety of organisational context controls, however, is that they do not offer by themselves any formal guarantee against inference attacks; further, quantified comparisons require estimations of their effectiveness, which can be difficult to make. This is the reason why we classify them as soft controls, whereas data controls that offer mathematical guarantees when combined with additional context controls can be deemed hard controls. This is not to say that hard controls should always be preferred to soft controls. The selection and combination of controls will depend upon a variety of factors and the specifics of each use case.

[53] L-diversity is an extension to k-anonymisation designed to mitigate inference attacks by preventing a significant fraction of each k-anonymous cohort from having similar values. This goal is to prevent attacks where, say, every member of the cohort possesses the same sensitive attribute as this would allow an attacker to make a sensitive inference despite being unable to single out an individual's record.

[54] T-closeness is similar to l-diversity except it requires that the distribution of values of a sensitive attribute, when restricted to any cohort, remains close in a certain formal sense to its distribution as observed across the whole dataset.

VI. How to Inform Legal Interpretation

With this background in mind, it is now possible to reassess anonymisation, de-identification, and aggregation provisions of privacy and data protection legislation and offer interpretative guidance.

The following two premises should now be taken for granted:

1. Both de-identified and aggregated data require context controls.
2. Techniques with formal mathematical guarantees should be preferred because they lend well to quantification of control effectiveness.

The legal standards are obviously different from the requirement that a formal guarantee against all four types of inferences should be achieved through the transformation process (ie, anonymisation under EU law or de-identification or aggregation under Californian law). Rather, it is expressed in terms of '*reasonable means likely to be used for re-identification*'[55] or '*reasonably linkable*.'[56] What this implies is that there is no requirement in the law that all four types of inferences should be mitigated through mathematical guarantees.

What has been debated, though, is whether all four types of inferences should be mitigated at all with K El Eman et al, for example, arguing that US healthcare law is only concerned with identity inferences.[57] K El Eman et al have thus been criticising the EU approach as described in ex-Art 29 WP's opinion on anonymisation techniques as being too restrictive.[58]

What seems clear, however, is that modern privacy and data protection laws such as the CCPA or GDPR are as a matter of principle concerned with a wider range of issues than the protection of the confidentiality of the identity of consumers or data subjects.[59] While this is made more explicit in the GDPR, which lists seven data protection principles, including data minimisation and fairness, and regulates profiling more strictly, it is also implicit in the definitions of consumer personal information as well as de-identified information and consumer information in the aggregate that are included in the CCPA.

This consideration should therefore have a direct impact upon the way the scope of these privacy and data protection laws is delineated and should require an initial assessment of all four types of inferences. However, depending upon the use case at stake, some forms of inferences (eg, relational inference and participation inference) could certainly be addressed through context controls only. Importantly, this is not suggesting that anonymised longitudinal studies are not possible anymore because relational inference would have to be

[55] GDPR, Recital 26.

[56] Cal. Civ. Code §§ 1798.140(a) and (h).

[57] K El Emam and C Alvarez, 'A Critical Appraisal of the Article 29 Working Party Opinion 05/2014 on Data Anonymization Techniques' (2015 5(1) *International Data Privacy Law* 73–87, https://doi.org/10.1093/idpl/ipu033.

[58] El Emam and Alvarez, 'A Critical Appraisal of the Article 29' (n 57).

[59] Decisions like the ones rendered under older privacy laws such as the 1984 Cable Communications Privacy Act or the 1998 Video Privacy Protection Act in the US should arguably not be of great help to understand key concepts. See, eg, *Pruitt v Comcast Cable Holdings, LLC*, 100 F. App'x 713, 716 (10th Cir. 2004). Compare with *In re Hulu Privacy Litig.*, No C 11-03764 LB, 2014 WL 1724344, at 10-11 (N.D. Cal. Apr. 28, 2014); *Yershov v Gannett Satellite*

mitigated through data controls. It is suggesting, on the contrary, that at a minimum strict attribute-based and role-based access control combined with an obligation not to further share the data should be in place to mitigate against participation, relational, and attribute inferences. For instance, while it may be technically possible to construct effective data controls for certain limited types of longitudinal analysis, doing so is burdensome and rigid in the sense that assumptions about the nature of the final result could hinder exploratory analysis and require vast expertise. Given the rigidity and costs of these data controls, it could make sense to enforce narrow permitted purposes via context controls and impose legal obligations upon data users.

Studies that require high-fidelity access to sensitive data clearly present an obstacle even for sophisticated data controls. This often includes longitudinal studies where lengthy records tend to leak relatively large amounts of private data. The inability to effectively leverage data controls without hampering utility must be offset by increasing context controls to compensate for increased overall risk.

As a matter of practice, sufficient context controls should be deployed such that the residual risk of accidental or malicious behaviour of individuals is viewed as mitigated. When the fidelity of sensitive data must be kept intact, it may be necessary to exclude all but a few, high trusted individuals in a controlled environment. This, in principle, is a valid approach to reduce the overall risk. By implementing strict access control, low-trust individuals are kept away from the data. Further, by enforcing obligations not to further share the data, the risk due to the accidental or malicious misbehaviour of the data user can be mitigated, as long as monitoring and auditing of data usage are enabled.

What is more, given the start of the art of data and context controls, privacy and data protection regulations should in fact be converging on matters relating to anonymisation, de-identification and aggregation, despite their differences in wording.

As a matter of principle two routes should lead to alleviating restrictions imposed by privacy or data protection laws:

1. **Local anonymisation:** the process by which the ability to make inferences from event-level data is limited for the release context in which the attacker operates.
2. **Aggregate anonymisation:** the process by which the ability to make inferences from aggregate data is limited for the release context in which the attacker operates.

Pseudonymisation, understood as the masking of direct identifiers, should be seen as a valuable security and data minimisation measure and constitutes the first step of any anonymisation process. Notably, both the GDPR and CCPA refer to pseudonymisation and distinguish it from de-identification, aggregation or anonymisation.[60]

[60] See GDPR, Art 4 and Recital 26; Cal. Civ. Code §1798.140(r).

The two anonymisation routes mentioned above appear compatible with the spirit of the law of modern privacy and data protection,[61] although their letter can appear problematic, which is the case for the CCPA in particular. Older statutes, such as the HIPAA and its Safe Harbor provision would need to be modernised.[62] HIPAA Safe Harbor appears to be particularly problematic, as it only mandates one data control and no context controls: the removal of 18 identifiers (ie, a partial version of control 4). This has been acknowledged by the US National Committee on Vital and Health Statistics in 2017, which recommends restricting downstream use of data even when complying with the HIPAA Safe Harbor[63]. The HIPAA expert determination provision[64] is, however, more flexible and makes it possible to consider both data and context controls as a means to address the four types of inferences aforementioned.

CCPA section 1798.140(h) governing de-identified data (which should correspond to the local anonymisation route) is interesting in that it lists key context controls and business processes that prohibit re-identification and prevent inadvertent release of personal information on top of technical safeguards. These business processes should however be also relevant for aggregated and synthetic data, which is not something that is expressly acknowledged by section 1798.140 governing aggregate consumer information (ie, what we conceive as aggregate anonymisation).[65] A reasoning by analogy would thus be needed for the interpretation of section 1798.140. What is important to bear in mind is that these business processes should always be assessed in the light of the inference-mitigation goal they seek to achieve. Of note, as synthetic data is a subset of aggregate data, it should be captured by section 1978.140 as it stands.

CCPA section 1798.145(a)(5) is confusing and could lead to unsatisfactory results if interpreted without due consideration of section 1798.140. Section 1798.145(a)(5) seems to suggest that both de-identified data and information in the aggregate can circulate freely, without context controls. Yet, the very definition of de-identified data includes context controls and as explained above, we suggest that the definition of information in the aggregate should implicitly comprise context controls as well.

GDPR Recital 26 is less specific than the CCPA but is worth pointing to for one reason: the mention that controls should be monitored over time. Both anonymisation routes seem available under the GDPR. The fact that pseudonymised data is considered to be personal data does not mean, as a matter of principle, that local anonymisation is not possible. As aforementioned, local anonymisation implies the treatment of both direct and indirect identifiers, whereas pseudonymisation is usually understood as a technique that masks direct identifiers only.

[61] After all, the CCPA includes within its definition of personal information inferences. See Cal. Civ. Code §1798.140(o)(1)(K).

[62] 45 CFR § 164.514(b)(2).

[63] National Committee on Vital and Health Statistics, 'Re: Recommendations on De-Identification of Protected Health Information under HIPAA,' 2017, www.ncvhs.hhs.gov/wp-content/uploads/2013/12/2017-Ltr-Privacy-DeIdentification-Feb-23-Final-w-sig.pdf. The National Committee makes recommendations related to the de-identification of protected health information under HIPAA and suggests controlling downstream use by the recipient of the data through access control. ('For example, covered entities and business associates might consider intended uses or the security and access controls used by recipients of a particular de-identified data set, in addition to considering the attributes of the data set.')

[64] 45 CFR § 164.514(b)(1). De-identifying data through the expert determination route requires the intervention of an expert who will be asked to determine 'that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information' as per 45 CFR § 164.514(b)(1)(i).

[65] What is more, the first prong ('Has implemented technical safeguards that prohibit reidentification of the consumer to whom the information may pertain') is not properly drafted in that it seems to suggest that technical safeguards alone are sufficient to prevent re-identification. The verb '*prohibit*' should therefore be understood in the sense of '*appropriately mitigate*.' In addition, the exclusion of publicly available information from the definition of personal information (Cal. Civ. Code § 1798.140(o)(2)) should not prevent the consideration of additional information to determine the re-identification risk level.

The GDPR is only one piece of the EU jigsaw and other legislations such as the Clinical Trial Regulation[66] should also be taken into account to make sense of the EU framework. Under Article 43 of the Clinical Trial Regulation, sponsors are required to submit annually 'a report on the safety of each investigational medicinal product used in a clinical trial for which it is the sponsor'.[67, 68] What is more, this report 'shall only contain aggregate and anonymised data'.[69] This would thus seem to exclude the local anonymisation route as we defined it above.

This exclusion makes sense in a clinical trial context. This is because it should be clear that longitudinal data can only be transformed into anonymised data within a closed environment. Therefore, public release could only happen if the aggregation route is chosen and if differential privacy is implemented and combined with additional context controls (ie, query monitoring and query termination or privacy budget).

It is true that Art 29 WP Anonymisation Techniques Opinion is not always easy to reconcile with a risk-based approach to anonymisation. While the Opinion has the merit that it is comprehensive, in that it covers a wide range of techniques and goes beyond the concern of identity disclosure or identity inference, it does include statements which are not compatible with a risk-based approach, such as the requirement that raw data should be destroyed to pursue anonymisation.[70] With this said, the Opinion does not suggest that it is not possible to mitigate re-identification risks through a combination of data and context controls. In any case, national regulators are not always aligned with the most contentious part of the Opinion[71] and other sector-specific regulatory authorities at the European level have issued guidance suggesting that a risk-based approach to de-identification remains a valid option, even after the Art 29 WP Anonymisation Techniques Opinion. Notably, this is the case of the European Medical Agency.[72]

This chapter does not suggest that because privacy or data protection restrictions are eliminated as a result of the combination of data and context controls, no harm could ever be caused to individuals. It has been demonstrated that in a machine learning context collective harm can be caused to individuals whose data has not been used to generate the models.[73] It is therefore crucial that an ethical impact assessment always be conducted. This essentially boils down to documenting and assessing model assumptions and limitations within the context of use cases illustrating how the model will work once deployed.

[66] Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC Text with EEA relevance [2014] OJ L158/1–76.

[67] The Court of Justice of the European Union (CJEU) actually recognised in Case C-582/14 *Breyer Patrick Breyer v Bundesrepublik Deutschland* 19 October 2016 ECLI:EU:C:2016:779 (*in fine*) the importance of context controls when applying the identifiability test.

[68] Clinical Trial Regulation, Art 43(1).

[69] ibid Art 43(3).

[70] 'Thus, it is critical to understand that when a data controller does not delete the original (identifiable) data at event-level, and the data controller hands over part of this dataset (for example after removal or masking of identifiable data), the resulting dataset is still personal data.' Art 29 WP Anonymisation Techniques (n 10) 9.

[71] See, eg, ICO Anonymisation Code of Practice, 21. While the Code was released in 2012 it was not amended after 2014 ('This does not mean though, that effective anonymisation through pseudonymisation becomes impossible. The Information Commissioner recognises that some forms of research, for example longitudinal studies, can only take place where different pieces of data can be linked reliably to the same individual.'). See also 58–59.

[72] European Medical Agency, External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use, 15 October 2018, EMA/90915/2016 Version 1.4. ('since in order to achieve a maximum usefulness of the data published, it is unlikely that for clinical reports all three criteria can be fulfilled [Possibility to single out an individual, Possibility to link records relating to an individual, Whether information can be inferred concerning an individual] by any anonymisation solution, it is EMA's view that a thorough evaluation of the risk of re-identification needs to be performed').

[73] See, eg, Ellen W McGinnis et al, 'Giving Voice to Vulnerable Children: Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood' (2019) 23(6) *IEEE Journal of Biomedical and Health Informatics* 2294–2301, https://doi.org/10.1109/JBHI.2019.2913590; Sophie Stalla-Bourdillon et al, 'Warning Signs — The Future of Privacy and Security in the Age of Machine Learning' (Future of Privacy Forum and Immuta Whitepaper, 2019), www.immuta.com/warning-signs-the-future-of-privacy-and-security-in-the-age-of-machine-learning/.

Notably, the GDPR framework is superior to the CCPA framework at least in relation to one key process-based requirement: assuming the data analytics process initiates with personal data, which is a sensible assumption to make as training data is likely to be transformed or protected in steps as explained above, an impact assessment of collective harms should always be included in the data protection impact assessment in situations of high risks. This is because data controllers are required to assess the impact of high risks to all 'the rights and freedoms of natural persons'.[74]

With this said, it should not be forgotten that as individual inferences (ie, output data produced when a model is applied upon an individual's data when making an individual decision) are personal data,[75] collective harm eventually leads to individual harm and as such will be captured by the privacy or data protection framework at this later point in time.

[74] GDPR, Art 40.

[75] CCPA defines inferences as 'the derivation of information, data, assumptions, or conclusions from facts, evidence, or another source of information or data.' Cal. Civ. Code § 1798.140(m) and includes individual inferences within the list of personal data. Cal. Civ. Code § 1798.140(o)(K): 'Inferences drawn from any of the information identified in this subdivision to create a profile about a consumer reflecting the consumer's preferences, characteristics, psychological trends, predispositions, behaviour, attitudes, intelligence, abilities, and aptitudes.' The same should be true with the EU data protection framework, even though the CJCE held in JEU, Joined Cases C-141/12 and C-372/12, *YS v Minister voor Immigratie, Integratie en Asiel and Minister voor Immigratie, Integratie en Asiel v M and S*, 17 July 2014, ECLI:EU:C:2014:2081 that a legal decision is not personal data. The distinction to draw is between the model, ie the reasoning, and the output, ie the inference. See, however, Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI,' 2018, https://doi.org/10.7916/D8-G10S-KA92. This is not to say that models can never remember personal information, in particular when controls have not been put in place during the training phase. See Veale, Binns, and Edwards, 'Algorithms That Remember' (n 14).

VII. Conclusion

To conclude, refining and implementing the inference model as developed in the literature on differential privacy we have demonstrated that there is no reason to think that aggregated or synthetic data are inherently safe.

Both aggregation and synthesisation should not be considered as effective mitigating strategies without the addition of data and context controls. As a consequence, privacy and data protection laws should not carve out exceptions for aggregate or synthetic data without requesting the combination of data and context controls. Data controls are controls that directly transform the data, while context controls affect the data environment and reduce the range of actions available to a data user.

We argue that two routes can lead to anonymisation: local anonymisation and aggregate anonymisation, kept at the event or individual level or aggregated, depending on the characteristics of the use case. In both cases, anonymisation can only be achieved if four types of inferences are taken into account and addressed either through data or context controls with a view to make inferences from data limited for the release context. We show that such an approach does not necessarily mean that it becomes impossible to anonymise longitudinal data.

Finally, we make the case that interpretation of anonymisation, de-identification or aggregation legal provisions should be converging and offer guidance to interpret recent provisions such as the CCPA section 1798.140 and GDPR Recital 26. We suggest that a risk-based assessment provides an objective measure of anonymisation approaches and has the potential to be replicable as long as assumptions related to attack methodologies hold. It should inform future interpretation of both the GDPR and CCPA. Nonetheless, given the confusing language used in provisions dealing with de-identification or anonymisation, in particular in the CCPA, a more detailed specification of the rules would be worth exploring. Guidance on attack methodologies would also prove extremely useful for organisations acting as data controllers.

Ultimately, what should be clear is that the binary dichotomy personal data/anonymised data is misleading for two reasons at least: it is not enough to look at the data to legally characterise the data and it is the potential for inferences which should drive the anonymisation approach, rather than actual inferences.

Acknowledgements

The authors are grateful to Dara Hallinan, Ronald Leenes, Serge Gutwirth, Paul de Hert; as well as the reviewers of this chapter for their insightful comments.

References

Abowd, John M 'Disclosure Avoidance and the 2018 Census Test: Release of the Source Code.' The United States Census Bureau. www.census.gov/newsroom/blogs/research-matters/2019/06/disclosure_avoidance.html (last accessed 4 March 2020).

Bellovin, Steven M, Preetam K. Dutta, and Nathan Reitinger. 'Privacy and Synthetic Datasets' (2019) 22(1) *Stanford Technology Law Review*.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 'Calibrating Noise to Sensitivity in Private Data Analysis' in Shai Halevi and Tal Rabin (eds) *Theory of Cryptography*, 3876:265–84. (Springer Berlin Heidelberg, 2006). https://doi.org/10.1007/11681878_14.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 'Calibrating Noise to Sensitivity in Private Data Analysis' (2017) 7(3) *Journal of Privacy and Confidentiality* 17–51. https://doi.org/10.29012/jpc.v7i3.405.

Dwork, Cynthia, and Aaron Roth. 'The Algorithmic Foundations of Differential Privacy' (2013) 9(3-4) *Foundation and Trends® in Theoretical Computer Science* 211–407. https://doi.org/10.1561/0400000042.

El Emam, K, and C Alvarez. 'A Critical Appraisal of the Article 29 Working Party Opinion 05/2014 on Data Anonymization Techniques' (2015) 5(1) *International Data Privacy Law* 73–87. https://doi.org/10.1093/idpl/ipu033.

El Emam, Khaled. *Guide to the De-Identification of Personal Health Information* 1st edn. (Auerbach Publications, 2013) https://doi.org/10.1201/b14764.

El Emam, Khaled, and Bradley Malin. *Appendix B: Concepts and Methods for De-Identifying Clinical Trial Data. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk* (National Academies Press (US), 2015) www.ncbi.nlm.nih.gov/books/NBK285994/.

Finch, Kelsey. 'A Visual Guide to Practical Data De-Identification.' Future of Privacy Forum. https://fpf.org/2016/04/25/a-visual-guide-to-practical-data-de-identification/ (last accessed 4 March 2020).

Fredrikson, Matthew, Eric Lantz, Somesh Jha, Simon Lin, David Page, and Thomas Ristenpart. 'Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing.' In Proceedings of the 23rd USENIX Conference on Security Symposium, 17–32. SEC'14. San Diego, CA: USENIX Association, 2014.

Garfinkel, Simson L. 'De-Identification of Personal Information.' National Institute of Standards and Technology, October 2015. https://doi.org/10.6028/NIST.IR.8053.

Gödel, Kurt. 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I.' [1931] 38(1) *Monatshefte für Mathematik und Physik* 173–98. https://doi.org/10.1007/BF01700692.

Green, Matthew. 'What Is Differential Privacy?' A Few Thoughts on Cryptographic Engineering (blog), 15 June 2016. https://blog.cryptographyengineering.com/2016/06/15/what-is-differential-privacy/.

Hu, Runshan, Sophie Stalla-Bourdillon, Mu Yang, Valeria Schiavo, and Vladimiro Sassone. 'Bridging Policy, Regulation and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR.' In *Data Protection and Privacy: The Age of Intelligent Machines* edited by Ronald Leenes Rosamunde van Brakel, Serge Gutwirth and Paul De Hert, 115-142. Oxford: Hart Publishing, 2017.

Kasiviswanathan, Shiva Prasad, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 'What Can We Learn Privately?' (2011) 40(3) *SIAM Journal on Computing 40* 793–826. https://doi.org/10.1137/090756090.

Mackey, Elaine, Mark Elliot, Kieron O'Hara, *The Anonymisation Decision-making Framework*, (UKAN Publications, 2016), https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf.

McGinnis, Ellen W, Steven P Anderau, Jessica Hruschak, Reed D Gurchiek, Nestor L Lopez-Duran, Kate Fitzgerald, Katherine L Rosenblum, Maria Muzik, and Ryan S McGinnis. 'Giving Voice to Vulnerable Children: Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood' (2019) 23(6) *IEEE Journal of Biomedical and Health Informatics* 2294–2301. https://doi.org/10.1109/JBHI.2019.2913590.

McMahan, Brendan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 'Communication-Efficient Learning of Deep Networks from Decentralized Data.' In Artificial Intelligence and Statistics, 1273–82, 2017. http://proceedings.mlr.press/v54/mcmahan17a.html.

McMahan, H Brendan, Galen Andrew, Ulfar Erlingsson, Steve Chien, Ilya Mironov, Nicolas Papernot, and Peter Kairouz. 'A General Approach to Adding Differential Privacy to Iterative Training Procedures.' ArXiv:1812.06210 [Cs, Stat], March 4, 2019. http://arxiv.org/abs/1812.06210.

National Committee on Vital and Health Statistics. 'Re: Recommendations on De-Identification of Protected Health Information under HIPAA,' 2017. www.ncvhs.hhs.gov/wp-content/uploads/2013/12/2017-Ltr-Privacy-DeIdentification-Feb-23-Final-w-sig.pdf.

Ohm, Paul. 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 *UCLA Law Review* 1701.

OneTrust DataGuidance Staff, and Future of Privacy Forum Staff. 'Comparing Privacy Laws: GDPR v. CCPA.' Accessed March 4, 2020. https://fpf.org/2019/12/18/comparing-privacy-laws-gdpr-v-ccpa/ (last accessed 4 March 2020).

Patki, Neha, Roy Wedge, and Kalyan Veeramachaneni. 'The Synthetic Data Vault.' In 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 399–410. Montreal, QC, Canada: IEEE, 2016. https://doi.org/10.1109/DSAA.2016.49.

Polonetsky, Jules, Omer Tene, and Kelsey Finch. 'Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification' (2016) 56(3) *Santa Clara Law Review* 593.

Samarati, Pierangela, and Latanya Sweeney. 'Protecting Privacy When Disclosing Information: K-Anonymity and Its Enforcement through Generalization and Suppression,' 1998, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.37.5829.

Staff, Ftc. 'Protecting Consumer Privacy in an Era of Rapid Change–A Proposed Framework for Businesses and Policymakers' [2011] *Journal of Privacy and Confidentiality*. https://doi.org/10.29012/jpc.v3i1.596.

Stalla-Bourdillon, Sophie, Brenda Long, Patrick Hall, and Andrew Burt. 'Warning Signs — The Future of Privacy and Security in the Age of Machine Learning.' Future of Privacy Forum and Immuta Whitepaper, 2019. www.immuta.com/warning-signs-the-future-of-privacy-and-security-in-the-age-of-machine-learning/.

Stalla-Bourdillon, Sophie, Alfred Rossi, and Gabriela Zanfir-Fortuna. 'Data Protection by Process: How to Operationalize Data Protection by Design,' 2019. https://fpf.org/2019/12/19/new-white-paper-provides-guidance-on-embedding-data-protection-principles-in-machine-learning/.

Valiant, LG 'A Theory of the Learnable.' Communications of the ACM 27, no 11 (November 5, 1984): 1134–42. https://doi.org/10.1145/1968.1972.

Veale, Michael, Reuben Binns, and Lilian Edwards. 'Algorithms That Remember: Model Inversion Attacks and Data Protection Law' (2018) 376(2133) *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 20180083. https://doi.org/10.1098/rsta.2018.0083.

Wachter, Sandra, and Brent Mittelstadt. 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI,' 2018. https://doi.org/10.7916/D8-G10S-KA92.

Warner, Stanley L 'Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias' (1965) 60(309) *Journal of the American Statistical Association* 63–69. https://doi.org/10.1080/01621459.1965.10480775.

IMMUTA™

115 Broad Street, 6th Floor, Boston, MA 02110 | immuta.com | (800) 655-0982

© 2020 Immuta, Inc. All rights reserved. 031521