Sustainable Privacy

Samuel M. Smith, Ph.D.

v0.2.0 2023/09/15

Abstract—Sustainable privacy takes a more holistic view of privacy protection. Conventional protection mechanisms such as K-anonymity-based de-identification are shown to be non-viable due to various attacks. The contextual linkability re-identification attack on selective disclosure is also described. Sustainable privacy requires, at the least, a more comprehensive approach that combines both technical and legal protections.

Index Terms—privacy, de-identification, re-identification, chain-link confidentiality, contextual linkability.

1 Introduction

Stated simply, sustainable privacy refers to the maintenance and preservation of personal data privacy over time. There are two sides to sustainable privacy. One side is how a person maintains their personal data privacy over time, and the other is how a holder of personal data respects and supports personal data privacy over time.

The difficulty of sustainable privacy is that any information gathering, usage, and sharing system is leaky. Information leaks out over time, thereby gradually eroding the privacy of any data being managed. Certainly, the sharing of data is inherently leaky, i.e., it is intentional and is, therefore, problematic when viewed from the perspective of sustainable privacy.

This report outlines the technical and legal issues surrounding sustainable privacy and guides better sustainable personal data privacy.

1.1 Three-Party Exploitation Model

Sustainable privacy is based on a three-party exploitation model. Fundamentally the goal is to protect the person (data subject) from exploitation via their personal data. In common usage, exploitation is selfishly taking advantage of someone to profit from them or otherwise benefit oneself. So, any unintended usage by any party is potentially exploitive. Intent is with respect to the person (data subject).

In this model, the 1st party is the person (data subject) of the original data. Their data is 1st party data. A 2nd party is the direct recipient of 1st party data as an intended recipient by the 1st party. A 3rd party is any other party who obtains or observes 1st party data but who is not the intended recipient. There are two main avenues of exploitation of 1st party data. These are any 2nd party who uses the data in any way not intended by the 1st party and any 3rd party who uses 1st party data. To clarify, any unintended (unpermissioned) use of 1st party data by any 2nd party is naturally exploitive.

Moreover, because a 3rd party is defined as an unintended recipient, any use of 1st party data by a 3rd party is likewise, by definition, exploitive.

Furthermore, 1st party data may be conveyed by one 2nd party to another 2nd party (i.e. shared) in a non-exploitive manner when such conveyance and eventual use by the other 2nd party is intended (permissioned) by the 1st party.

This model is diagrammed below.

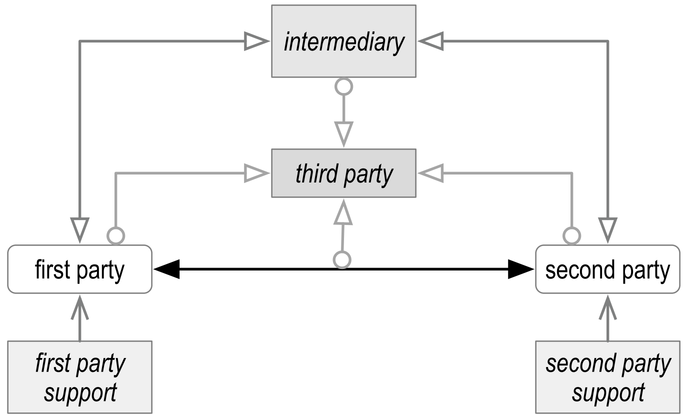

Figure 1.1. Three-Party Exploitation Model

The diagram above includes all the vectors by which data may leak and thereby erode the sustainable privacy of 1st party data. This report primarily focuses on how 2nd parties may either help or hinder sustainable privacy.

1.2 Data Privacy

Information or data privacy is defined is the relationship between the collection and dissemination of data, technology, the public expectation of privacy, contextual information norms, and the legal and political issues surrounding them. Data privacy is challenging since it attempts to allow the use of data by 2nd and 3rd parites while protecting personal (1st party) privacy preferences and personally identifiable information (PII). The fields of computer security, data security, and information security all design and use software, hardware, and human resources to address this issue [22]. This definition is consistent with privacy viewed from the perspective of 1st party data rights and the role of 2nd parties in the three-party exploitation model. The Trust over IP (ToIP) foundation's architecture specification phrases privacy protection as answering the question:

> *Privacy: will the expectations of each party with respect to the usage of shared information be honored by the other parties?[44]*

The primary mechanisms by which 2nd parties erode sustainable data privacy rights are as follows:

- Exploitive use of 1st party data by 2nd parties.
- Sharing of 1st party data by 2nd parties with 3rd parties either overtly or inadvertently via leakage.

The conventional best practices for managing personal data are often based on the principles of privacy by design and privacy by default (PbDD) [3].

Recent advances in technology, however, have made many conventional best practices for such PbDD policy insufficient for sustainable privacy. This report will explain why. One of those recognized best practices for PbDD is called data de-identification (or data anonymization).

The most common approach to de-identification is based on an approach called K-anonymity [23]. The basic idea is that a data set may be de-identified by deleting any attributes in a data set that contain personally identifying information (PII) such as name, address, and phone number. More sophisticated approaches also delete sufficiently identifying quasi-identifiers. A quasi-identifier is any subset of attributes that together are personally identifying. Indeed, various K-anonymity approaches to data de-identification have become the principal best practice for protecting the privacy of 1st party data when sharing with 3rd parties. The attraction of K-anonymity as a best practice is that it gives any 2nd party a safe harbor for sharing any and all 1st party data with any 3rd party for any purpose. Basically, if the data has been de-identified using K-anonymity, then by definition, it is no longer 1st party data. Unfortunately, K-anonymity as a de-identification technique is easy to defeat and, therefore, should no longer be considered an acceptable best practice [2; 4–6; 16; 18; 24; 25; 27; 36; 43].

There are two other mechanisms for de-identification that so far are considered practically infeasible to defeat and hence are suitable for sustainable privacy. These are differential privacy and synthetic data privacy [1; 10–15; 28; 30; 31; 41; 42; 46].

Given that K-anonymity is broken and viable alternatives to K-anonymity exist, the use of K-anonymity can no longer be considered a best practice for de-identification of data before sharing. The only allowable de-identification approach is either a differential privacy or synthetic data privacy-based mechanism.

The next sections explore in more detail the primary factors underlying sustainable privacy and provide guidance for best practices in achieving it.

2 Re-identification

Let's first consider the problem of de-identification and re-identification of data. It was long thought that de-identification or anonymization of data could provide privacy using a technique called K-anonymity [23]. Indeed the ease and pervasiveness by which de-identified data may be re-identified have resulted in the US FTC (Federal Trade Commission) issuing a warning that those who share de-identified databases and purport that merely through conventional de-identification that the privacy rights of the associated persons are protected may be in violation of the laws regulating the use and sharing of sensitive data [5; 40].

2.1 Linkage Re-identification Attacks

Recent research has shown that fully de-identified sparse datasets (using K-anonymity) can be merged to re-identify the data [2; 8; 9; 25; 27; 35]. In 2022, this was further extended to what is called a down-coding attack which enables the re-identification of data even when every field is a quasi-identifier, i.e., there is no personally identifying information in the dataset [4]. These are all examples of what are called linkage attacks [32]. In general, the vulnerability to linkage attacks inherent in conventional de-identification mechanisms means that using such for maintaining data set privacy is an exercise in risk management rather than a proviso of guaranteed protection [36].

2.2 Profiling Re-identification Attack

Online interactions between parties captured as interactions of anonymized members of a social graph using only time, duration, and type of interaction are enough to re-identify the majority of the members of the social graph using only a 2-hop interaction graph [6]. The latter is an example of a new type of attack called a profiling re-identification attack, where machine learning is used to re-identify based on de-identified behavioral data and not merely by linking de-identified database attributes (quasi-identifiers or non-identifiers).

2.3 Contextual Linkability Re-identification Attack

As mentioned above, any de-identified dataset, even when all attributes are quasi-identifiers, may be vulnerable to re-identification through various statistical correlation attacks, be they linkage or behavioral profiling-based correlation. In this section, we introduce a new correlation attack called contextual linkability that defeats the purported privacy protection mechanism commonly known as selective disclosure [17].

The selective disclosure, whether via Zero-Knowledge-Proof (ZKP) or not, of any 1st party data disclosed to a 2nd party may be potentially trivially exploitably correlatable via the re-identification correlation techniques described above when those techniques are applied to auxiliary data obtained at the time of presentation of the selectively disclosed attributes[47] . We call an attack that uses the statistical correlation of auxiliary attributes obtained from the context of a disclosure a contextual linkability re-identification attack. Essentially, it's the use of a set of non-selectively disclosed attributes (auxiliary data) obtained from the context of the disclosure that is sufficient to re-identify the discloser despite selective disclosure. Thus, a contextual linkability attack may trivially defeat the cryptographic unlinkability provided by selective disclosure mechanisms, including those that use ZKPs.

In essence, the vulnerability stems from the fact that the verifier may structure the context of the presentation so as to provide sufficient auxiliary data that the combination of contextual auxiliary data and selectively disclosed data is identifying. In other words, contextual linkability has the potential to create a data set of quasi-identifiers that may be combined with selectively disclosed attributes in such a way as to re-identify the associated subject of the selectively disclosed attributes. Thus ZKPs or other selective disclosure mechanisms by themselves may provide insufficient privacy protection. Any unconstrained selectively disclosed set of attributes is inherently re-identifiable unless the discloser takes care to both protect from contextual linkability and impose a contractual liability on any use of that data. Essentially bare selective disclosure implicitly grants the disclosee (verifier) a safe harbor to use, assimilate, and re-identify the disclosed data.

This limitation relegates the status of selective disclosure and/or ZKPs as privacy mechanisms to the narrow corner conditions where there is zero contextual linkability by the 2nd party disclosee (verifier) at the time of presentation. Because most presentation contexts are under the control of the 2nd party (verifier), the verifier needs merely to structure that context with enough quasi-identifier attributes (auxiliary data) to re-identify the presenter, which in turn would enable the 2nd party to link the de-identified presenter back to the issuer with the presentation details thereby defeating cryptographic unlinkability. This cryptographic unlinkability is the only unique reason to use a ZKP in the presentation in the first place. This does not mean that there are no corner conditions where the presentation context is sufficiently under the control of the presenter (1st) party such that the presenter can structure the context so as to prevent the verifier from correlating any other quasi-identifiers, but none of the standard use cases for cryptographic unlinkability satisfy that condition.

Correctly understood, selective disclosure is a naive form of K-anonymity performed by the discloser (presenter). The discloser is attempting to de-identify their own data. Unfortunately, such naive de-identified disclosure is not performed with any statistical insight into the ability of the verifier (receiver) to re-identify the selectively disclosed attributes given the contextual attributes that are also disclosed (inadvertently) at the time of presentation and under the control of the verifier.

In light of this vulnerability, many of the standard use cases for selective disclosure (with or without ZKPs) in verifiable credentials (VCs) are examples of anti-patterns for privacy protection [45]. This is because these standard use cases assume a presentation context that is under the control of the verifier, which means a malicious verifier can restructure that context to statistically guarantee correlation and defeat the selective disclosure with or without a ZKP.

For example, when a disclosure is made at the place of business of the receiver (verifier), the receiver may use readily available location data from mobile phone providers to re-identify based on a geo-fence query or readily available facial identification based on on-premise security camera footage. Similarly, if the presentation of a facial biometric is required, then the receiver may use facial recognition systems as well to re-identify the presenter. If payment is required, then the receiver may use credit card information to re-identify the presenter. Likewise, when the disclosure is made to the receiver's website, the receiver may use readily available IP source addressing and routing information to re-identify the presenter. Moreover, for social applications that employ bare repeated selective disclosures (even with ZKPs), the set of social interactions observable by the 2nd parties form a context that may be a correlatable social graph. Therefore a profiling reidentification attack that merely requires the time, duration, and type of interaction associated with each selective disclosure may reidentify the selective disclosures.

Any receiver (verifier) sophisticated enough to verify a ZKP selective disclosure presentation is sophisticated enough to use these readily available contextual re-identification techniques to defeat that selective disclosure.

Indeed, systems that use selective disclosure as an advertised privacy protection mechanism may result in a net decrease in privacy protection for users because of the false belief that selective disclosure alone is sufficient to protect the disclosure. This may increase re-identifiable disclosures that would not have happened otherwise. Users may be lulled into a false belief that because their selectively disclosed attributes do not include any personally identifying information, there is no need for them to impose any constraints on the use and/or sharing of their naively de-identified (selectively disclosed) attributes. Indeed, the receiver may surreptitiously induce such unconstrained disclosures by reinforcing the false belief that the de-identified attributes are not (easily) re-identifiable. For this reason, using a bare selective disclosure mechanism may be considered irresponsible by any organization that purports to use it for privacy protection.

The core defect of any K-anonymity-like approach, including selective disclosure, is that there is no apriori way to establish if any attribute is an identifier, quasi-identifier, or non-identifier. All attributes are potentially identifying attributes based on the available auxiliary data. This includes auxiliary data obtained from the context of the disclosure. As a result, bare (naive) selective disclosure alone provides no guarantee of privacy protection to the discloser. Thus de-identifying aggregated 1st party data, including de-identification via selective disclosure, provides no meaningful privacy protection.

This means that naive K-anonymity-based mechanisms such as selective disclosure, including ZKPs, that provide so-called cryptographic unlinkability may now be trivially linkable by statistical re-identification methods. Any selective disclosure is potentially ineffective unless performed within the confines of a contractually protected disclosure that imposes an incentive on the disclosee (verifier) to protect that disclosure (counter-incentive against the exploitation of that discloser).

3 Protection from Re-identification Attacks on 1st Party Data

There are two modes of data sharing with respect to 1st party data. The first is the sharing of non-aggregated 1st party data, and the second is the sharing of already aggregated 1st party data.

3.1 Non-aggregated 1st Party Data

Any non-aggregated 1st party data shared with a 2nd party may be easily re-identifiable because there is no herd privacy. The data is directly attributable to the one and only one 1st party. The act of sharing forms a sharing disclosure context that may be structured by the 2nd party to provide easily correlatable contextual auxiliary data that enables re-identification via a contextual linkage attack. Consequently, due to the ease of re-identification via a contextual linkage attack, the only practically viable protection against the 2nd-party correlation of non-aggregated 1st-party data is pre-disclosure contractual protection by imposing liability on the 2nd party disclosee and strict post-disclosure chain-link confidentiality on any downstream disclosees or other users of that data, including any later assimilation or aggregation [21]. Chain-link confidentiality can impose a requirement that 2nd parties may not aggregate 1st party data, de-identified or not (selectively disclosed or not), without the consent of the 1st party.

To reiterate, the only sustainable privacy protection mechanism of 1st party data disclosed to a 2nd party, even when selectively disclosed (via ZKP or not), is contractually protected disclosure via chain-link confidentiality.

3.2 Aggregated 1st Party Data

As mentioned above, aggregated 1st party data (multiple 1st parties) may be de-identified using one or both of two statistically provable de-identification techniques, namely differential privacy and/or synthetic data privacy.

3.2.1 Differential Privacy

When used properly, differential privacy takes a data set and corrupts the field values with statistic noise so that it is difficult or impossible to re-identify any individuals [10–13][14; 28; 46]. A major drawback of differential privacy is that the corrupting noise may change the statistics of the aggregated data such that certain types of critical inference are also corrupted, as in the case of data used for medical risk and efficacy [29].

3.2.2 Synthetic Data Privacy

Synthetic data privacy is based on a self-supervised machine learning technique that uses real data as the training set to synthesize a new data set with similar statistics to the training set but which synthetic data set provides resistance to re-identification of the real data records given the synthetic data [1; 7; 15; 30; 37; 41; 42]. Therefore, a caveat in generating synthetic data is that the degree of learning accuracy must be limited to not compromise re-identification resistance [31; 34].

3.2.3 Comparison

Synthetic data privacy is more computationally intensive than differential privacy but has the potential for much higher fidelity of its aggregated statistics to the real data used in its training set while maintaining a comparable level of re-identification resistance. Thus synthetic data may be more valuable than differentially privatized data. In both cases, care must be taken to ensure sufficient statistical un-correlatability (un-linkability) between the synthetic or differentially privatized data and the real data.

3.3 Protection via Management of the Time Value of Information

In general, privacy dissipates over time. This is because digital information is inherently leaky, and those leaks become more correlatable as the body of leaks grows over time. The diminishing exploitable time value of correlated information can balance this leakiness.

The primary exploitable time value of correlated information for data aggregators is that it can be used to predict behavior. Advertisers want to predict who will most likely be receptive to their marketing campaigns. The predictive accuracy of aggregated behavioral information of potential participants for any given market-related behavior is largely a function of the nearness in time of that market-related behavior when used to make the prediction. We can ascribe a time constant to a given market for the exploitable predictive potential of market-related behavior where information older than the time constant no longer has net predictive value in excess of the cost of aggregating it. Information that exceeds this time constant is considered stale because there is no longer any incentive to aggregate and correlate it. Therefore, in spite of the fact that privacy dissipates over time, the value of correlation also diminishes over time so that cost-effective privacy protection mechanisms can focus resources on near-term correlatability. This provides a sweet spot for sustainable privacy protection that is governed by the time constant of the time value of exploitable correlation. Likewise, the cost of privacy protection can be weighed against the cost due to the harm of exploitation. If the cost of protection exceeds the cost due to the harm of exploitation, then it's not worth protecting (i.e., it's counterproductive to the protector). If the cost of correlation in order to exploit exceeds the time value return of exploitation, then it's not worth exploiting (i.e., it's counterproductive to the exploiter).

One approach to managing this trade-off would be a protocol for exchanging or sharing information that granular partitions data-sharing contexts so that correlatability is also granularly partitioned. The Trust over IP (ToIP) is developing such a protocol called the Trust Spanning Protocol (TSP) [26; 39].

The TSP protocol enables one to control the exploitable time value of the correlatable information that can be leaked from a given context. Once a context has become leaky, however, a new isolated context can be created that restarts the clock on time-value correlatability. This provides a trade-space between the friction and cost of forming and maintaining contexts, the length of time before a given context becomes leaky, and the time constant of the exploitable value of leaked information. The cost of protection includes expensive one-time OOBA (Out-of-Band-Authentication) setups. If the leakiness of a given context is cost-effectively protectable beyond the time constant on the time value of exploitation, then new information is sustainably protectable indefinitely.

4 Legal and Policy Issues

In general, the idea that conventional de-identification by removing PII grants collectors of personal data a safe harbor with respect to protecting personal privacy data rights is an outdated notion. Simply stated, conventional anonymization does not work. All data collected about a person must be considered sensitive and must only be used by intended users for intended purposes. The practice of effectively *privacy-washing* the data via conventional de-identification in order to share it with no strings attached is a problematic practice. Instead, data sharing must be process-based and contextual, with strings attached [36]. Even when the newer, more advanced techniques of differential privacy or synthetic data privacy are employed to more effectively de-identify data sets, the sharing of personal data without the consent of the person is inherently exploitive.

One way to attach strings to the data that conveys the person’s intent for its usage is through chain-link confidentiality [21]. Chain-link confidentiality provides legal protection for privacy by leveraging the well-established framework of confidentiality law. This imbues the collection, holding, and sharing of personal data with explicit confidentiality protection. This may be considered a policy of confidentiality by design. With chain-link confidentiality, all personal data can have strings attached for all downstream users and uses of that data. Complimentary to explicit confidentiality protection is implicit confidentiality protection[19]. Implied confidentiality is a default policy that augments both confidentiality by design and privacy by design by leveraging confidentiality law. Together, explicit confidentiality and implicit confidentiality by default provide comprehensive privacy protection wrapped in the framework of confidentiality law. The emerging open ACDC standard includes support for chain-link confidentiality [38].

As an overarching principle, any privacy-respecting 2nd party should behave under a duty of loyalty to the preferences and expectations of the data privacy rights of 1st parties. This principle is succinctly referred to as data loyalty [20; 33]. The 2nd party data holder must act as a loyal agent of the 1st party. In other words, the 2nd party should be loyal to the best interests of any 1st party with respect to their data.

From a legal and policy perspective, a loyal data agent acts as a fiduciary on behalf of the data subject. By definition, a fiduciary must act in the best interests of its client, which is in direct opposition to what an exploiter of personal data would do. It is clear that many, if not most, 2nd parties act more like data exploiters than data fiduciaries.

Exploitation is not limited to selling data; it includes any form of data opportunism that benefits the 2nd party at the expense of a 1st party. Exploitation also includes using data for nudging and manipulation. A fiduciary is responsible for negligence when it allows exploitation by others or when it misses opportunities to use that data to benefit its clients according to the expressed best interests of those clients. From that perspective, a loyal data agent can be incentivized by being held accountable for mis-, mal-, and non-feasance with respect to this duty.

Acknowledgments

The author wishes to thank all those who provided helpful feedback.

Author

Samuel M. Smith, Ph.D., has a deep interest in decentralized identity and reputation systems as well as the authenticity, confidentiality, and privacy implications of the associated technology. Samuel received a Ph.D. in Electrical and Computer Engineering from Brigham Young University in 1991. He then spent 10 years at Florida Atlantic University, eventually reaching full professor status. He has performed pioneering research in automated reasoning, machine learning, and autonomous vehicle systems as well as decentralized identity and reputation. He has over 100 refereed publications in these areas and was the principal investigator on numerous federally funded research projects. Dr. Smith has been an active participant in open standards development for network and identity protocols. He is also a serial entrepreneur.

References

[1] "13 Tools for Synthetic Data Generation to Train Machine Learning Models," Geekflare,
https://geekflare.com/synthetic-data-generation-tools/

[2] "33 Bits," 33 Bits of Entropy,
https://33bits.wordpress.com/about/

[3] Cavoukian, A., "Privacy by Design The 7 Foundational Principles,"
https://iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf

[4] Cohen, A., "Attacks on Deidentification's Defenses," vol. 31st USENIX Security Symposium, no. USENIX Security 22, pp. 1469-1486, 2022
https://www.usenix.org/conference/usenixsecurity22/presentation/cohen

[5] Cohen, K., "Location, health, and other sensitive information: FTC committed to fully enforcing the law against illegal use and sharing of highly sensitive data," FTC Business Blog, 2022/07/11
https://www.ftc.gov/business-guidance/blog/2022/07/location-health-and-other-sensitive-information-ftc-committed-fully-enforcing-law-against-illegal

[6] Crețu, A.-M., Monti, F., Marrone, S. et al., "Interaction data are identifiable even across long periods of time," Nature Communications, vol. 13, no. 1, pp. 313, 2022
https://www.nature.com/articles/s41467-021-27714-6

[7] Dankar, F. K. and Ibrahim, M., “Fake it till you make it: Guidelines for effective synthetic data generation,” Applied Sciences, vol. 11, no. 5, pp. 2158, 2021

https://www.mdpi.com/2076-3417/11/5/2158/pdf

[8] De Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M. and Blondel, V. D., “Unique in the crowd: The privacy bounds of human mobility,” Scientific reports, vol. 3, no. 1, pp. 1-5, 2013

https://www.nature.com/articles/srep01376

[9] De Montjoye, Y.-A., Radaelli, L., Singh, V. K. and Pentland, A. “. S., “Unique in the shopping mall: On the reidentifiability of credit card metadata,” Science, vol. 347, no. 6221, pp. 536-539, 2015

https://dspace.mit.edu/bitstream/handle/1721.1/130329/Unique in the shopping mall- On the reidentifiability of credit card metadata.pdf?sequence=1

[10] “Differential Privacy,” Wikipedia,

https://en.wikipedia.org/wiki/Differential_privacy

[11] “Differential Privacy,” Harvard University Privacy Tools Project,

https://privacytools.seas.harvard.edu/differential-privacy

[12] “Differential Privacy,” Apple.com,

https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf

[13] Dwork, C., “Differential privacy: A survey of results,” vol. International conference on theory and applications of models of computation, pp. 1-19, 2008

https://web.cs.ucdavis.edu/~franklin/ecs289/2010/dwork_2008.pdf

[14] Dwork, C. and Roth, A., “The algorithmic foundations of differential privacy,” Foundations and Trends® in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211-407, 2014

https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf

[15] Eastwood, B., “What is synthetic data — and how can it help you competitively?,” MIT Management Sloan School, 2023/01/23

https://mitsloan.mit.edu/ideas-made-to-matter/what-synthetic-data-and-how-can-it-help-you-competitively

[16] “First-Party Data: What It Is, How to Use It, and Why It Matters Now More Than Ever,” LiveRamp, 2022/10/12

https://liveramp.com/explaining-first-party-data/

[17] Flamini, A., Ranise, S., Sciarretta, G. et al., “A First Appraisal of Cryptographic Mechanisms for the Selective Disclosure of Verifiable Credentials,”

https://www.researchgate.net/publication/370897881_A_First_Appraisal_of_Cryptographic_Mechanisms_for_the_Selective_Disclosure_of_Verifiable_Credentials/link/6467fdba66b4cb4f73c1b46e/download

[18] Fredrikson, M., Jha, S. and Ristenpart, T., “Model inversion attacks that exploit confidence information and basic countermeasures,” vol. Proceedings of the 22nd ACM SIGSAC conference on computer and communications security, pp. 1322-1333, 2015

https://dl.acm.org/doi/pdf/10.1145/2810103.2813677

[19] Hartzog, W., “Reviving Implied Confidentiality,” Ind. LJ, 2014

https://www.repository.law.indiana.edu/cgi/viewcontent.cgi?article=11108&context=ilj

[20] Hartzog, W. and Richards, N., “The Surprising Virtues of Data Loyalty,” Emory LJ, 2021

https://scholarship.law.bu.edu/cgi/viewcontent.cgi?article=4057&context=faculty_scholarship

[21] Hartzog, W., “Chain-link confidentiality,” Ga. L. Rev., vol. 46, pp. 657, 2011

https://digitalcommons.law.uga.edu/cgi/viewcontent.cgi?article=1323&context=glr

[22] “Information privacy,” Wikipedia,

https://en.wikipedia.org/wiki/Information_privacy

[23] “K-anonymity,” Wikipedia,

https://en.wikipedia.org/wiki/K-anonymity

[24] "Level Up Your Loyalty Marketing With Zero-Party Data," Forrester, 2022/09/14
https://www.forrester.com/blogs/level-up-your-loyalty-marketing-with-zero-party-data/

[25] "Lie of Anonymous Data," TechCrunch, 2019/07/04
https://techcrunch.com/2019/07/24/researchers-spotlight-the-lie-of-anonymous-data/

[26] "Mid-year Progress Report on the ToIP Trust Spanning Protocol," ToIP, 2023/08/31
https://trustoverip.org/blog/2023/08/31/mid-year-progress-report-on-the-toip-trust-spanning-protocol/

[27] Narayanan, A., Huey, J. and Felten, E. W., "A precautionary approach to big data privacy," Data protection on the move: Current developments in ICT and privacy/data protection, pp. 357-385, 2016
http://ndl.ethernet.edu.et/bitstream/123456789/66793/1/352.pdf#page=375

[28] Nguyen, A., "Understanding Differential Privacy," Towards Data Science, 2019/06/30
https://towardsdatascience.com/understanding-differential-privacy-85ce191e198a

[29] Oberski, D. L. and Kreuter, F., "Differential privacy and social science: An urgent puzzle," Harvard Data Science Review, vol. 2, no. 1, pp. 1-21, 2020

[30] Patki, N., Wedge, R. and Veeramachaneni, K., "The synthetic data vault," vol. 2016 IEEE International Conference on Data Science and Advanced Analytics, no. DSAA, pp. 399-410, 2016
https://ieeexplore.ieee.org/abstract/document/7796926

[31] Platzer, M. and Reutterer, T., "Holdout-based empirical assessment of mixed-type synthetic data," Frontiers in big Data, vol. 4, pp. 679939, 2021
https://www.frontiersin.org/articles/10.3389/fdata.2021.679939/full

[32] Powar, J. and Beresford, A. R., "SoK: Managing risks of linkage attacks on data privacy," Proceedings on Privacy Enhancing Technologies, vol. 2, pp. 97-116, 2023
https://petsymposium.org/popets/2023/popets-2023-0043.pdf

[33] Richards, N. and Hartzog, W., "A Duty of Loyalty for Privacy Law," Wash. UL Rev., 2021
https://scholarship.law.bu.edu/cgi/viewcontent.cgi?article=4055&context=faculty_scholarship

[34] Riemann, R., "Synthetic Data," European Data Protection Supervisor,
https://edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en

[35] Rocher, L., Hendrickx, J. M. and De Montjoye, Y.-A., "Estimating the success of re-identifications in incomplete datasets using generative models," Nature communications, vol. 10, no. 1, pp. 1-9, 2019
https://www.nature.com/articles/s41467-019-10933-3

[36] Rubinstein, I. S. and Hartzog, W., "Anonymization and risk," Wash. L. Rev., 2016
https://digitalcommons.law.uw.edu/cgi/viewcontent.cgi?article=4948&context=wlr

[37] "Self-Supervised Learning," Wikipedia,
https://en.wikipedia.org/wiki/Self-supervised_learning

[38] Smith, S. M., "Authentic Chained Data Containers (ACDC)," ToIP Foundation,
https://github.com/trustoverip/tswg-acdc-specification

[39] Smith, S. M., "Secure Privacy, Authenticity, and Confidentiality (SPAC)," 2023
https://github.com/SmithSamuelM/Papers/blob/master/whitepapers/SPAC_Message.md

[40] Solove, D. J. and Hartzog, W., "The FTC and the new common law of privacy," Colum. L. Rev., 2014
https://scholarship.law.bu.edu/cgi/viewcontent.cgi?article=4036&context=faculty_scholarship

[41] "Synthetic Data Vault Tools," Datacebo,
https://sdv.dev

[42] "The real promise of synthetic data," MIT News, 2020/10/16
https://news.mit.edu/2020/real-promise-synthetic-data-1016

[43] "These startups are helping online marketers get around Apple's privacy changes," FastCompany,

https://www.fastcompany.com/90773053/online-marketing-privacy-tracking-apple-facebook

[44] "Trust over IP (ToIP) Technology Architecture Specification," ToIP,
https://github.com/trustoverip/TechArch/blob/main/spec.md#61-design-goals

[45] "Verifiable Credentials," Wikipedia,
https://en.wikipedia.org/wiki/Verifiable_credentials

[46] Wood, A., Altman, M., Bembenek, A. et al., "Differential privacy: A primer for a non-technical audience," Vand. J. Ent. & Tech. L., vol. 21, pp. 209, 2018
https://dash.harvard.edu/bitstream/handle/1/38323292/4_Wood_Final.pdf?sequence=1

[47] "Zero-knowledge proof," Wikipedia,
https://en.wikipedia.org/wiki/Zero-knowledge_proof