A New Approach to Utility-based Privacy Preserving in Data Publishing

Yılmaz Vural
Department of Computer Engineering
Hacettepe University
06800, Beytepe, Ankara, TURKEY
yvural@hacettepe.edu.tr

Murat Aydos
Department of Computer Engineering
Hacettepe University
06800, Beytepe, Ankara, TURKEY
maydos@hacettepe.edu.tr

***Abstract*— A fundamental problem in privacy-preserving data publishing is how to make the right trade-off between privacy risks and data utility. Anonymization techniques are the primary tools used to reduce privacy risks. Generalization with full suppression technique is commonly used for anonymization. Due to the fact that fully suppressed outlier records are not generalized, data utility is negatively affected in the process. In this study, a new approach is proposed by reducing the number of outlier records in order to increase the data utility. In the proposed model, k-anonymity and l-diversity privacy models are used together to reduce the privacy risks. The Avarege Equivalence Class Size is used in measuring the data utility. According to the experimental results, the data utility is increased while keeping the delicate balance between privacy risks and data usefulness.**

***Index Terms*— data anonymization; data publishing; data privacy; data utility; privacy models.**

I. INTRODUCTION

In today's digital age, online services collect large amounts of data about their users and activities. The diversity and the volume of the collected data is rapidly increasing. These data are shared with researchers or institutions according to their requirements. There are personal and sensitive information in the shared data causing privacy risks. Data privacy is a difficult problem that tries to find the best trade-off solution between privacy and utility [1, 2]. Privacy models are used to solve the particular privacy problem.

Data anonymization is needed to implement the privacy model, which constitutes the process of de-identifying sensitive data while preserving its format and data type [3]. The combined use of generalization and suppression is the main method for anonymization operations [4]. This process is called recoding [5]. It replaces a value with another, less specific but semantically consistent value to hide certain details.

Data to be anonymized are typically released in the form of a microdata table [6]. It is defined as a set of attributes that can be classified as identifier (ID), quasi identifier (QID), sensitive attribute (SA) and non-sensitive attribute (NSA). ID attributes can be used to identify the data owner. ID attributes are associated with a high risk of re-identification. They have to be removed from the data set. Typical examples are names or social security numbers. QID attributes that can be linked with external information to re-identify the data owner. They have to be transformed by an anonymization technique. Typical examples are gender, date of birth and ZIP codes. SA attributes represent sensitive information for the data owner. They might be of interest to an attacker. In the case of a disclosure, the attacker could attack the privacy. These sensitive and private attributes should be kept unmodified for data utility but may be subject to further constraints. Typical examples are health diagnoses and personal salaries. NSA attributes are generally not considered to be sensitive by the data owner and the release of them is harmless [7]. Insensitive attributes are not associated with any privacy risk. They will be kept unmodified. Equivalent Classes (EC) are records with the same QID attribute values [8]. The whole record set in the EC cannot be distinguished from each other. Outlier Equivalent Classes (OEC) consist of records with no data utility [9].

A fundamental problem in privacy-preserving data publishing is how to make the right trade-off between privacy risks and data utility [10]. There are privacy models that try to find the best trade-off between privacy risks and data utility. In this study, the important privacy models are reviewed. *(1) K-anonymity*, which was proposed by Sweeney[11], requires that each record of a released table cannot be identified with a probability higher than 1/k, which means that each record is indistinguishable from at least the other k-1 records. *(2) L-diversity*, which was proposed by Machanavajjhala et al. [12, 13], requires that each EC has at least l well-represented value for each sensitive attribute. *(3) T-closeness*, which was proposed by Li et al. [14], requires that the distribution of a sensitive attribute in any EC is close to the distribution of the attribute in the overall table.

There are two important parameters in solving the problem of data privacy [15]. One of them is measuring data utility. The important data utility metrics are reviewed in this work. The first is the Discernibility Metric (DM), which was proposed by Bayardo et al. [16]. It is based on an EC and introduces a penalty for the suppressed entries. The second one is the Average Equivalence Class Size (AECS), which was proposed by LeFevre et al. [17]. It is based on an EC and it is measured according to the number of equivalent classes. The third one is the Information Loss (IL), which was proposed by Iyengar et al.

[18]. It summarizes the coverage of the domain of an attribute. The other parameter is estimating privacy risk. Attacker models can be used to estimate the privacy risks. There are three different attacker models for measurement of privacy risks [19]. The first attacker model is the prosecutor model. Here it is assumed that the attacker already knows that the data set contains data about the respondent. The second one is the journalist model, which assumes that the attacker does not have background information about the respondent. The third one is the marketer model. It assumes that the attacker is not interested in re-identifying a specific individual but that s/he aims at attacking a larger number of individuals.

In summary, we investigated the impact of the OEC on the data utility and the privacy risks. The previous studies have shown that generalization with suppression leads to a significant increase in data utility [9]. However, it is clear that there is still room for improvement since these methods make no use of the OEC by completely ruling them out. The OEC includes outlier records that have been fully suppressed resulting in no data utility. In this study, we showed that using OEC in the anonymization process could increase the data utility. According to the findings obtained from our study, the proposed model improved the data utility compared to the previous results while keeping the privacy risk levels on the same amount.

II. Materials and Methods

In this section, we propose a new utility-based method and explain the experimental setup used in our evaluation.

A. Method

According to our model, to increase data utility without compromising privacy, Eq. 1 is applied to the dataset which included microdata.

$$T^*_{dataset} = \rho\text{-Gain}\ (T_{dataset}) \quad (1)$$

T is used to denote the original data set. T* is used to denote the anonymized data set. ρ is used to denote the number of iterations. Table-I gives the steps of the proposed model.

TABLE I. Steps of Model

Layers	Steps	Process
L_1 (Preparation Layer)	1.1	Classify attributes in the $T_{dataset}$ (ID, QID, SA, NSA)
	1.2	Remove ID attributes from the $T_{dataset}$
	1.3	Add GAIN (auxiliary attribute) to $T_{dataset}$
	1.4	Set GAIN=0 all records in $T_{dataset}$
	1.5	Create QID generalization hierarchies
	1.6	Set ρ'=0
	1.7	Copy all record from $T_{dataset}$ to $T'_{dataset}$
L_2 (Anonymiza-tion Layer)	2.1	Select coding method for $T_{dataset}$
	2.2	Select data utility metric for $T_{dataset}$
	2.3	Select k, l, t, for privacy models
	2.4	Input the ρ value
	2.5	If ρ' = ρ then go to **Step 3.4**
	2.6	Anonymize $T_{dataset}$ with privacy models
	2.7	Set GAIN=1 for all records in outlier EC
	2.8	Move all records from $T_{dataset}$ to $T^*_{dataset}$ where GAIN=0
L_3 (Gain Layer)	3.1	De-anonymized all records in $T_{dataset}$ use $T'_{dataset}$
	3.2	Set GAIN=0 all records in $T_{dataset}$
	3.3	ρ '= ρ' + 1 and go to **Step 2.5**
	3.4	Move all records from $T_{dataset}$ to $T^*_{dataset}$ where GAIN=0
	3.5	Remove GAIN (auxiliary attribute) attribute in $T^*_{dataset}$
	3.6	Copy all records from $T'_{dataset}$ to $T_{dataset}$ and delete $T'_{dataset}$
L_4 (Publishing Layer)	4.1	Publish $T^*_{dataset}$

B. Experimental Evaluations

We performed two kinds of experiments to evaluate our solution. First, we compared our model with related work in terms of data utility. In these experiments, we used suppression limits of 100% in order to find actual size of OEC. In the second set of experiments, we showed that our model did not cause any negative impact on the privacy risk estimates.

We used well-known models, which are k-anonymity and l-diversity together, to reduce privacy risks. We chose $k = 5$ and $l = 3$, which have been proposed in the literature [9, 20, 21]. According to our model, the inclusion of the OEC in the process has significantly increased the data utility. We measured data utility with AECS, which is a model based on the equivalence class sizes.

We have used two real-world datasets, most of which have been used in previous studies on data privacy. The datasets included a DEMOGRAPHICS [22] dataset and an ADULT dataset [23]. Specifically, the ADULT dataset is de-facto standard because it is often used in previous studies.

The datasets were prepared for the anonymization process. In this context, we removed the missing values and selected one SA and seven QID attributes of each dataset. The classification of attributes on these datasets are shown in Table II.

TABLE II. Classification of Attributes

Dataset	SA	QID (Height of generalization hierarchies)
ADULT	salary-class	age (5), education (4), marital-status (3), native-country (3), race (2), salary-class (2), sex (2), social class (3)
DEMOGRAPHICS	Income	age (5), gender (2), race (2), ethnic (2), education (3), marital-status (3), poverty (2)

We used ARX, which is a comprehensive open-source software for anonymizing data, to evaluate our model [24]. Experiments were performed on a desktop computer running ARX 3.5.1 with a quad-core 2.6 GHz Intel Core i7 CPU running a 64-bit Oracle JVM.

An example of OEC, which is the key point of this study, is given in Fig. 1. The OEC includes outlier records that have been fully suppressed resulting in no data utility. There is an inverse proportion between the data utility and the privacy.This is when one value increases as the other value decreases.

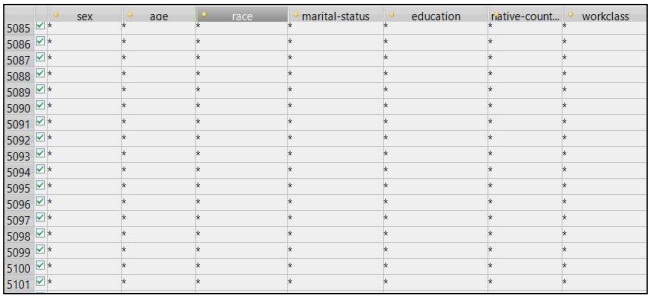

Fig. 1. An example of OEC

As shown in Fig.1, the privacy is the highest while the data utility is zero. For privacy preserving data publishing, which is a balance problem, this is an undesirable situation. In the generalization with suppression process, to increase the data utility, the number of outlier records has to decrease.

C. Re-Coding and Solution Space

In the experimental evaluations, we used the global recoding model with suppression as a coding method. Global recoding means that the same transformation is applied to all entries in the dataset. Suppression means that a complete record is replaced with any symbol (e.g., *). An example of the coding method used in this paper is shown in Table III.

TABLE III. (A) A RAW TABLE. (B) A TWO-ANONYMITY VIEW BY GLOBAL RECODING WITH SUPPRESSION.

Id	Age	Gender	Zip	Income
1	39	Female	06100	60K
2	43	Female	06100	40K
3	45	Male	34100	70K
4	48	Male	34100	45K
5	55	Female	46100	60K
6	58	Female	46100	50K

(a)

Age	Gender	Zip	Income
[39-44]	*	061**	60K
[39-44]	*	061**	40K
[45-49]	*	341**	70K
[45-49]	*	341**	45K
[54-59]	*	461**	60K
[54-59]	*	461**	50K

(b)

When global recoding with suppression is used as a coding model, it is possible to model the solution space as a generalization lattice, which is a partially ordered set of all possible combinations of generalization levels of each attribute [9]. In this study, we have used lattices that could be visualized with Hasse diagrams [25]. An example of a generalization lattice for the sex and zip attributes is shown in Fig. 2.

A generalization lattice is a partially ordered set of all possible combinations of generalization levels of each attribute. A node represents a transformation and defines generalization levels for QID attributes. Data utility is the highest in Level-0 (original dataset), while data privacy is the highest in Level-3 (full suppressed dataset).

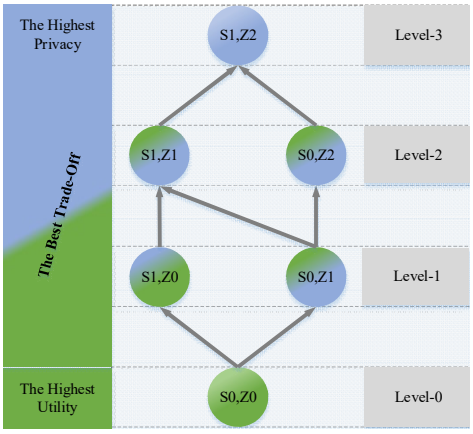

Fig. 2. Solution space for the S(ex) and Z(ip) attributes

III. RESULTS

In this section we present the experimental results and their analysis. In the first place, we present the result of data utility. In the second, the results of privacy risk estimates were exposed and analyzed.

A. Data Utility Results

According to our model, to increase data utility without compromising privacy, Eq. 1 was applied to the ADULT and DEMOGRAPHICS datasets, respectively. In our experimental evaluations, we measured the data utility with AECS, which is a model based on equivalence class sizes. As a result of our work, we set the iteration value (ρ) to 2. When ρ=2 is selected for the ADULTSET, Eq. 2 can be applied to it.

$$T^{*}_{ADULT} = 2\text{-Gain}\,(T_{ADULT}) \qquad (2)$$

For the first iteration (ρ'=0), the ADULT utility results are shown in Fig. 3. There are 13181 outlier records in the OEC, which includes 43.7% of the total data.

Measure	Including outliers	Excluding outliers
Average class size	25.47466 (0.08446%)	14.35418 (0.08453%)
Maximal class size	13181 (43.70068%)	80 (0.47111%)
Minimal class size	5 (0.01658%)	5 (0.02944%)
Number of classes	1184	1183
Number of records	30162	16981 (56.29932%)
Suppressed records	13181 (43.70068%)	0

Fig. 3. ADULT utility results (ρ'=0, k=5, l=2)

Similarly, when ρ=2 is selected for the DEMOGRAPHICS dataset, Eq. 3 can be applied to it.

$$T^{*}_{DEMOGRAPHICS} = 2\text{-Gain}\,(T_{DEMOGRAPHICS}) \qquad (3)$$

For the first iteration (ρ'=0), the DEMOGRAPHICS utility results are shown in Fig. 4. There are 11985 outlier records in the OEC, which includes 17.4% of the total data.

Measure	Including outliers	Excluding outliers
Average class size	9.50062 (0.01386%)	7.84105 (0.01386%)
Maximal class size	11985 (17.47951%)	19 (0.03358%)
Minimal class size	5 (0.00729%)	5 (0.00884%)
Number of classes	7217	7216
Number of records	68566	56581 (82.52049%)
Suppressed records	11985 (17.47951%)	0

Fig. 4. DEMOGRAPHICS utility results (ρ'=0, k=5, l=2)

For the second iteration (ρ'=1), the ADULT utility results are shown in Fig. 5. There are 7796 outlier records in the OEC, which includes 59,1 % of the total data.

Measure	Including outliers	Excluding outliers
Average class size	27.46042 (0.20833%)	11.24217 (0.20877%)
Maximal class size	7796 (59.14574%)	80 (1.48561%)
Minimal class size	5 (0.03793%)	5 (0.09285%)
Number of classes	480	479
Number of records	13181	5385 (40.85426%)
Suppressed records	7796 (59.14574%)	0

Fig. 5. ADULT utility results (ρ'=1, k=5, l=2)

For the second iteration (ρ'=1), the DEMOGRAPHICS utility results are shown in Fig. 6. There are 3849 outlier records in the OEC, which includes 32,1 % of the total data.

Measure	Including outliers	Excluding outliers
Average class size	14.50969 (0.12107%)	9.86182 (0.12121%)
Maximal class size	3849 (32.11514%)	23 (0.28269%)
Minimal class size	5 (0.04172%)	5 (0.06146%)
Number of classes	826	825
Number of records	11985	8136 (67.88486%)
Suppressed records	3849 (32.11514%)	0

Fig. 6. DEMOGRAPHICS utility results (ρ'=1, k=5, l=2)

For the third iteration (ρ'=2), the ADULT utility results are shown in Fig. 7. There are 6060 outlier records in the OEC, which includes 77,7 % of the total data.

Measure	Including outliers	Excluding outliers
Average class size	65.51261 (0.84034%)	14.71186 (0.84746%)
Maximal class size	6060 (77.73217%)	72 (4.14747%)
Minimal class size	5 (0.06414%)	5 (0.28802%)
Number of classes	119	118
Number of records	7796	1736 (22.26783%)
Suppressed records	6060 (77.73217%)	0

Fig. 7. ADULT utility results (ρ'=2, k=5, l=2)

For the third iteration (ρ'=2), the DEMOGRAPHICS utility results are shown in Fig. 7. There are 2423 outlier records in the OEC, which correspond to 62,9 % of the total data.

Measure	Including outliers	Excluding outliers
Average class size	55.78261 (1.44928%)	20.97059 (1.47059%)
Maximal class size	2423 (62.95142%)	40 (2.80505%)
Minimal class size	6 (0.15588%)	6 (0.42076%)
Number of classes	69	68
Number of records	3849	1426 (37.04858%)
Suppressed records	2423 (62.95142%)	0

Fig. 8. DEMOGRAPHICS utility results (ρ'=2, k=5, l=2)

The data utility comparison for each dataset is given in the following table IV.

TABLE IV. UTILITY SUMMARY (A) ADULTS. (B) DEMOGRAPHICS

ADULT	AECS	Number of Classes	Number of Records / Outlier Records
Before Our Model	14,35	1183	30162 /13181
After Our Model	13,54	1780	30162 / 6060

(a)

DEMOGRAPHICS	AECS	Number of Classes	Number of Records / Outlier Records
Before Our Model	7,84	7216	68566 /11985
After Our Model	7,37	8109	65866 / 6060

(b)

When our model was applied to the datasets, it was observed that the number of outlier records was decreased, resulting in an increase of the data utility.

B. Privacy Risk Estimate Results

The risk estimate for the ADULT data set according to the Prosecutor model is given in Fig. 9.

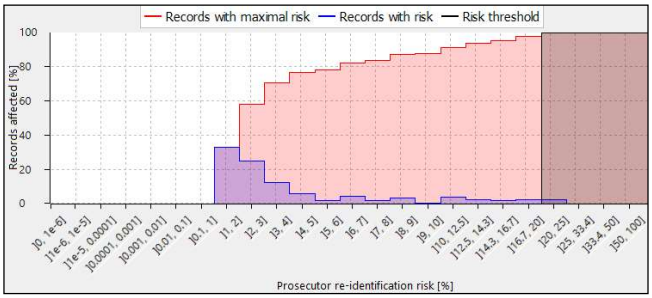

Fig. 9. Prosecutor re-identification risk for ADULTS

The re-identification risk was related to the size of the equivalence class. The highest risk estimation is about 20% for the ADULT dataset.

The risk estimate for the DEMOGRAPHICS dataset according to the Prosecutor model is given in Fig. 10.

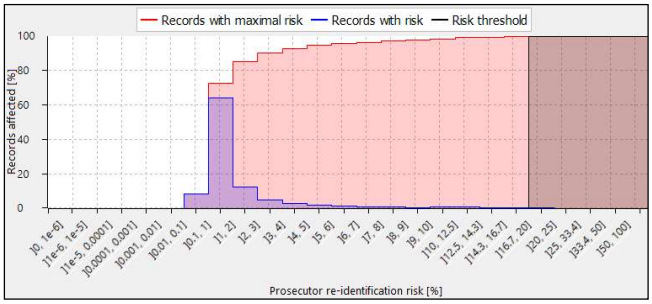

Fig. 10. Prosecutor re-identification risk for DEMOGRAPHICS

The maximum privacy risk of the DEMOGRAPHICS dataset is estimated at around 20% similar to the ADULT dataset. The privacy risk summary for the ADULT dataset is given in Fig. 11.

ADULT	ρ'=1	ρ'=2	ρ'=3
Measure	Value [%]	Value [%]	Value [%]
Lowest prosecutor risk	1.25%	1.25%	1.38889%
Records affected by lowest risk	0.47111%	1.48561%	4.14747%
Average prosecutor risk	6.96661%	8.89508%	6.79724%
Highest prosecutor risk	20%	20%	20%
Records affected by highest risk	4.24003%	4.27112%	2.88018%
Estimated prosecutor risk	20%	20%	20%
Estimated journalist risk	20%	20%	20%
Estimated marketer risk	6.96661%	8.89508%	6.79724%

Fig. 11. ADULT dataset re-identification risk estimates

In Fig. 11, three iterations are shown. It was observed that the privacy was preserved. The privacy risk summary for the DEMOGRAPHICS dataset is given in Fig. 12. In this figure, three privacy evaluation results are shown. It was seen that privacy was preserved. When our model is applied to the DEMOGRAPHICS dataset, the privacy risks are reduced.

DEMOGRAPHICS	ρ'=1	ρ'=2	ρ'=3
Measure	Value [%]	Value [%]	Value [%]
Lowest prosecutor risk	5.26316%	4.34783%	2.5%
Records affected by lowest risk	0.03358%	0.28269%	2.80505%
Average prosecutor risk	12.7534%	10.14012%	4.76858%
Highest prosecutor risk	20%	20%	16.66667%
Records affected by highest risk	8.81038%	3.13422%	0.42076%
Estimated prosecutor risk	20%	20%	16.66667%
Estimated journalist risk	20%	20%	16.66667%
Estimated marketer risk	12.7534%	10.14012%	4.76858%

Fig. 12. DEMOGRAPHICS dataset re-identification risk estimates

According to utility results, when our model applied to datasets, it was observed that the data utility is increased. Subsequently, privacy (re-identification) risks were not negatively affected by this process. In summary, the experimental results show that, our model increased data utility without compromising privacy.

IV. Conclusions

In this study, the problem of privacy preserving micro data publishing (PPDP) was addressed. We focused on data utility enhancement while maintaining a balance between data utility and privacy risk. In the previous studies, it has been shown that using the generalization with suppression technique increases the data utility. Disadvantages of the previous studies are the presence of outlier records that do not have data utility. For this reason, outlier records were included in the anonymization process for the first time in this study and the effects of this new approach were observed on data utility.

In our approach, we have used *k-anonymity* and *l-diversity* models together to preserve privacy, while the outlier records have been included in the process to enhance the data utility. Our proposed model was tested for data utility and privacy risks with two real world datasets. Our experimental evaluation results for data utility and privacy risks were compared to the previous studies. According to the results, we observed that the use of the OEC has a positive effect on the data utility, while causing no-negative impact on privacy risks. One of the main contributions of our work is to enhance data utility using the OEC.

This work can be further extended by working on different datasets with different data utility metrics to find the best trade-off between privacy risks and data utility. In addition, studies should be done to find the optimal number of iterations (ρ). Another important issue for further research is the effect of privacy parameters, data types, utility and risk metrics on the occurrence of outlier records.

References

[1] P. Samarati, "Protecting Respondents' Identities in Microdata Release," *IEEE Trans. on Knowl. and Data Eng.,* vol. 13, no. 6, pp. 1010-1027, 2001.

[2] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing," *ACM Computing Surveys,* vol. 42, no. 4, pp. 1-53, 2010.

[3] B. Raghunathan, "Overview of Data Anonymization," in *Complete Book of Data Anonymization: From Planning to Implementation* T. a. Francis, Ed. Florida/U.S.A: Auerbach Publications, 2013, pp. 1-13.

[4] Y. Xu, T. Ma, M. Tang, and W. Tian, "A survey of privacy preserving data publishing using generalization and suppression," *Applied Mathematics & Information Sciences,* vol. 8, no. 3, pp. 1103-116, 2014.

[5] G. Ghinita, P. Kalnis, and Y. Tao, "Anonymous publication of sensitive transactional data," *IEEE Transactions on Knowledge and Data Engineering,* vol. 23, no. 2, pp. 161-174, 2011.

[6] M. Petković and W. Jonker, "Privacy and Security Issues in a Digital World," in *Security, Privacy, and Trust in Modern Data Management*, M. Petković and W. Jonker, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 3-10.

[7] W.-Y. Lin, D.-C. Yang, and J.-T. Wang, "Privacy preserving data anonymization of spontaneous ADE reporting system dataset," *BMC Medical Informatics and Decision Making,* journal article vol. 16, no. 1, p. 58, 2016.

[8] I. Ozalp, M. E. Gursoy, M. E. Nergiz, and Y. Saygin, "Privacy-Preserving Publishing of Hierarchical Data," *ACM Trans. Priv. Secur.,* vol. 19, no. 3, pp. 1-29, 2016.

[9] F. Kohlmayer, F. Prasser, and K. A. Kuhn, "The cost of quality: Implementing generalization and suppression for anonymizing biomedical data with minimal information loss," *Journal of biomedical informatics,* vol. 58, pp. 37-48, 2015.

[10] T. Li and N. Li, "On the tradeoff between privacy and utility in data publishing," presented at the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France, 2009.

[11] L. Sweeney, "k-anonymity: a model for protecting privacy," *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.,* vol. 10, no. 5, pp. 557-570, 2002.

[12] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "L-diversity: privacy beyond k-anonymity," in *22nd International Conference on Data Engineering (ICDE'06)*, 2006, pp. 24-24.

[13] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Trans Knowl Discov Data,* vol. 1, no. 1, pp. 3-15, 2007.

[14] N. Li, T. Li, and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," in *2007 IEEE 23rd International Conference on Data Engineering*, 2007, pp. 106-115.

[15] B.-C. Chen, D. Kifer, K. LeFevre, and A. Machanavajjhala, "Privacy-Preserving Data Publishing," *Foundations and Trends® in Databases,* vol. 2, no. 1–2, pp. 1-167, 2009.

[16] R. J. Bayardo and R. Agrawal, "Data privacy through optimal k-anonymization," in *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on*, 2005, pp. 217-228: IEEE.

[17] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Mondrian multidimensional k-anonymity," in *Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on*, 2006, pp. 25-25: IEEE.

[18] V. S. Iyengar, "Transforming data to satisfy privacy constraints," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2002, pp. 279-288: ACM.

[19] K. El Emam, *Guide to the de-identification of personal health information*. CRC Press, 2013.

[20] F. Prasser and F. Kohlmayer, "Putting statistical disclosure control into practice: The ARX data anonymization tool," in *Medical Data Privacy Handbook*: Springer, 2015, pp. 111-148.

[21] F. Prasser, R. Bild, J. Eicher, H. Spengler, F. Kohlmayer, and K. A. Kuhn, "Lightning: Utility-Driven Anonymization of High-Dimensional Data," *Transactions on Data Privacy,* vol. 9, no. 2, pp. 161-185, 2016.

[22] Q. Gong. (2015). *Partition and Anatomize Anonymization* Available: https://github.com/qiyuangong/PAA/blob/master/data/demographics.csv

[23] M. Lichman. (2013, 30.03). *{UCI} Machine Learning Repository*. Available: http://archive.ics.uci.edu/ml

[24] F. Prasser, F. Kohlmayer, R. Lautenschläger, and K. A. Kuhn, "Arx-a comprehensive tool for anonymizing biomedical data," in *AMIA Annual Symposium Proceedings*, 2014, vol. 2014, p. 984: American Medical Informatics Association.

[25] R. Brüggemann and G. P. Patil, "Partial Order and Hasse Diagrams," in *Ranking and Prioritization for Multi-indicator Systems: Introduction to Partial Order Applications*New York, NY: Springer New York, 2011, pp. 13-23.