Integrating Differential Privacy in Modern Database Curriculum

Jeff Miller
School of Computing & Analytics
Northern Kentucky University
Highland Heights, Kentucky
millerj109@nku.edu

Ankur Chattopadhyay
School of Computing & Analytics
Northern Kentucky University
Highland Heights, Kentucky
chattopada1@nku.edu

***Abstract*—College students in computer science and information technology majors have traditionally been required to take database design and processing classes, as well as data analytics classes in both business and technology settings-- as part of the required curriculum. As outlined conceptually by The Joint ACM/AIS IS2020 task force, there is and will continue to be a great demand for students who have the skills to query for data, both in analytical tasks, as well as programming, statistics, administrative and managerial tasks. The information technology sector, focusing on the workplace and the application of information technology skills in business, would greatly benefit from new modern privacy concepts that have evolved recently as a result of the trends in big data, social media, IoT and business analytics. Cloud database, as well as traditional storage systems for data are subject to General Data Protection Regulation (GDPR) which requires entities such as businesses, companies, and organizations of all types who store data to take reasonable measures to protect a subject's data and privacy against data loss or exposure. This paper explores the research available in the subject of data privacy as applicable to database programming curriculum in information technology. It explores the ever-important use of structured query language in relation to data privacy. In addition, this paper outlines the concepts and definitions of differential privacy, used to obfuscate sensitive data collected and stored in modern relational database schemes. It seeks to provide knowledge for integrating the concept of differential privacy and closely related PETs into the traditional business model-based relational database management system class at the undergraduate level**

Keywords—Big Data, Cloud Data, GDPR, Privacy, Structured Query Language, IoT, Information Technology Curriculum, Statistics, Business Analytics, IS2010, IS2020, The Joint ACM/AIS IS2020 Task Force, Differential Privacy, Privacy Enhancing Technologies

I. Introduction and Motivation

In recently researched, relevant literature comprised of articles, websites, surveys, and presentations; privacy principals in regard to databases seems to have begun in 1977 when Tore Dalenius articulated a desideratum for statistical databases: nothing about an individual should be learnable from the database that cannot be learned without access to the database [19]. Since that time, privacy slowly evolved into a permanent aspect of Information Systems. Current guidelines for information technology standards exist in the form of IS2020, and before that, IS2010. Created by the Joint ACM/AIS Taskforce, their primary function in relation to curriculum is to help Information Systems programs produce competent and confident entry-level graduates well-suited to workplace responsibilities or further studies of Information Systems [5]. In order to keep up with the pace of technology, the model curriculum should be flexible and adaptable to most Information Systems programs [5]. In a recent survey of IS2010 and IS2020 publications, privacy as a concept is hardly addressed in IS2010, yet is mentioned in several sections of IS2020. DP or 'differential privacy' is part of a larger category of PET or 'privacy enhancing technologies', under the categorization of data obfuscation tools. PETs can be divided into four categories: data obfuscation, encrypted data processing, federated and distributed analytics and data accountability tools [12]. There are two natural models for privacy mechanisms: interactive and noninteractive. In the non-interactive setting the data collector, a trusted entity, publishes a "sanitized" version of the collected data; the literature uses terms such as "anonymization" and "de-identification". Traditionally, sanitization employs techniques such as data perturbation and sub-sampling, as well as removing well-known identifiers such as names, birthdates, and social security numbers. It may also include releasing various types of synopses and statistics. In the interactive setting the data collector, again trusted, provides an interface through which users may pose queries about the data, and get (possibly noisy) answers [18].

Data obfuscation tools include zero-knowledge proofs (ZKP), differential privacy, synthetic data, and anonymization and pseudonymization tools. These tools increase privacy protections by altering the data, by adding "noise" or by removing identifying details. Obfuscating data enables privacy-preserving machine learning and allows information verification (e.g., age verification) without requiring sensitive data disclosure. Data obfuscation tools can leak information if not implemented carefully however. Anonymized data for instance can be re-identified with the help of data analytics and complementary data sets [12].

Within a modern curriculum, differential privacy is relevant as a PET because it provides data subjects with some protection of deniability in cases where someone attempts to re-identify released data. Noise introduced into the dataset should not alter any large-scale analysis but makes any individual data less reliable and protective for the data subjects. Policy makers may need to provide guidance about the amount of noise that must be introduced to protect the privacy of data subjects [12].

II. Project Goals, Research Questions and Paper Overview

This research study, as presented in the following sections, seeks to answer the following research questions:

2024 IEEE Integrated STEM Education Conference (ISEC) | 979-8-3503-5280-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISEC61299.2024.10664891

- How thoroughly is data privacy discussed in IS2010 & IS2020? What differences exist between the existing IS2010 & IS2020 recommendations regarding data privacy?

- How does GDPR relate to modern relational database systems? Which types of data are pertinent to DP and covered by GDPR?

- What forms of DP exist? Which forms of DP relate to relational database systems? Is it possible to integrate DP into an undergraduate level database class?

- What are PPDM, PPDP, and PPGP? How does PPDM, PPDP, and PPGP relate to data privacy and the collection of data in relational database systems?

- How does SQL in general relate to DP? What are noteworthy iterations of SQL that have been developed to address privacy concerns?

- What kind of lecture information and hands-on work would be suitable to introduce to the undergraduate database curriculum?

In the next sections, we discuss the related works and research done on and around our curriculum, privacy-related, and SQL-based topics of this paper. Privacy needs to be addressed in the undergraduate database curriculum as demonstrated by the implications of exposing personal data. Furthermore, personally identifiable information is of increasing importance and the privacy of this data should be a priority in our database programming curriculum.

III. Discussions on Related Works and Research Background

The following section A-C on IS2010 and IS2020 provides insight into a discussion on the role of privacy and its relevancy in the curriculum. Sections D-E illustrate the research background deemed relevant to the creation of discussion and lecture material for use in an undergraduate database class.

A. IS2010 and IS2020

The IS2002 guidelines that preceded IS2010 were widely accepted and have been influential as the basis for the accreditation of undergraduate programs at a typical University. The Joint ACM/AIS IS2010 Curriculum guidelines only mentioned "privacy" in the following areas:

- 2010.1 - Foundations of Information Systems, pp. 391-393
- 2010.7 - IS Strategy, Management and Acquisition, pp. 402
- Elective Course IT Security and Risk Management, pp. 412
- Security and Privacy - Figure A4.2A pp. 422
- Privacy - Figure A4.2B pp. 423

The existence of privacy as a concept and term in general, is only used on the above listed pages in The Joint ACM/AIS IS2010 Curriculum guidelines. The IS2020 report is reflective of the latest IS discipline guidelines used in 2024. This IS2020 report constitutes the combined effort of numerous individuals and has been designed to reflect the interests of many more faculty and practitioners [6]. Some implications and consequences resulting from this revolution will be controversial and even negative, threatening the basic rights of citizens, and creating hazards for societies. To deal with potential adverse consequences of the information explosion, governments and other regulators are developing new legislation and standards. As an example, such regulations can deal with the collection and use of personal data (e.g., the EU General Data Protection Regulation). Societal and regulatory changes relating to privacy and ethical issues also suggest the need for updates to curriculum recommendations for the knowledge of rules, ethics, and regulations affecting IS [6].

IS2020 lists the information consolidated in Table 1 to illustrate the changes from IS2010.

TABLE I. PRIVACY RELATED COMPETENCY IN IS2020 CURRICULUM

Related Competency in IS2020	
Knowledge Elements	***Skill Level***
Types of information privacy threats	2 - Understand
Consequences of information privacy violations	2 - Understand
Technologies and solutions for information privacy	2 - Understand
Fair information practices and privacy policies	2 - Understand
Government information privacy regulations	2 - Understand
Analyze the importance of social media privacy and security	4 - Analyze
Privacy trade-offs and risks in the social context	4 - Analyze
Database creation	6 - Create
Skills to create a web application using front- and back-end development along with incorporating database functionality (CRUD)	6 - Create
Website database encryption and decryption	6 - Create

As you can see from Table 1, privacy and secure computing is an increasingly important competency as people become more reliant on technology. Although two security courses were previously included as electives in IS2010 recommendations; the past decade has seen an increased rise in security and privacy violations and technological developments to address them; thus, confirming the increasing importance of this topic for modern organizations and for the IS profession [6]. These recommendations are not only useful to information systems students, but also useful for related areas; a finance and accounting major, with specialization on auditing, may find databases, technology infrastructures, and computing security useful. A marketing major is more likely to be interested in big data analytics, IS use and ethics, or application development [6].

In conclusion, to further illustrate the importance of privacy and the potential links to differential privacy in the context of database, I will state that the expected database outcomes for database include the following competencies on pp. 102 are:

- Query the relational model
- Design relational databases
- Programming database systems using functions and triggers
- Secure a database
- Compare tradeoffs of different concurrency modes
- Develop non-relational models [6]

Similar outcomes will be addressed later in this paper, as we look to the future of database in the undergraduate curriculum in the form of a current undergraduate syllabus at a major research institution, Southern Illinois University.

B. GDPR

In response to the rapid loss of data privacy over the past decade, governments have begun developing new privacy regulations (e.g. the General Data Protection Regulation (GDPR) in the European Union). These regulations recognize the im- portance of privacy and attempt to specify how it must be protected [2]. Today's information systems collect and process vast amounts of data, and the majority of it flows into databases (relational or otherwise). These database systems are specifically designed to collect, store, and query data, and have been optimized for that task. If we would like to enable an analysis of sensitive data with differential privacy, it is logical to develop techniques that work for database systems, because that's where the private data is [8]. Private data is the most pertinent to GDPR. Today's information systems collect and process vast amounts of data, and the majority of it flows into databases (relational or otherwise). These database systems are specifically designed to collect, store, and query data, and have been optimized for that task. If we would like to enable an analysis of sensitive data with differential privacy, it is logical to develop techniques that work for database systems, because that's where the private data is [8]. Traditionally, Data Management has focused on data persisting in organizations, usually in relational databases. Such data assets support the core business processes of the organization and form the basis for business applications. Increasingly, organizations also process ever larger volumes of data that emerge from expansive digitalization (web traffic, social media, and sensed sources). Regardless of the source and type of data, the fundamental questions and concerns of this realm remain the same: How to gather, organize, curate, and process data to help run an organization or extract actionable information to increase effectiveness. The Data/Information competency realm comprises one required area (Data and Information Management) and two elective areas (Data and Business Analytics; Data and Information Visualization) [6]. Where does differential privacy fit in the new world of regulated privacy? Unfortunately, it depends on your interpretation of the law [2].

C. Differential Privacy

Differential privacy (DP) [20] is a well-known and mathematical definition-based privacy protection model. It is mostly used for privacy protection in interactive settings of the PPDP. It protects the privacy of the user by adding noise to the original user's data and it does not make assumptions about the intruder scenarios. The DP belongs to the semantic class of privacy models, and it yields superior privacy protection in PPDP compared to the syntactic privacy models. Considering the effectiveness of the DP model, U.S. Census Bureau is planning to use the DP in their 2020 census, and all future data products [21]. Differential privacy is a simple mathematical definition that indicates when publishing results or data sets can be considered 'private' in a specific sense. The term, its definition, and many of the modern techniques associated with it, were Invented by theoretical computer scientists Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith (see Dwork & Roth, 2014, for key references). These researchers took a step back from the field initiated by Dalenius and Fellegi and rebuilt its foundations on a rigorous definition that could be used to protect data [3].

It has been reported in the literature that DP provides a mathematically provable guarantee on privacy preservation against many privacy attacks such as differencing, linkage, and reconstruction attacks [4]. Knowledge discovery from data, or KDD – typically refers to the process composed of the following sequence of steps: data cleaning; data integration; data selection; data transformation; data mining; pattern evaluation; and knowledge presentation [9].

The fundamental goal of differential privacy is to prevent the leakage of private information to an adversary. DP achieves this goal by providing a provable guarantee, a generic bound on privacy leakage that makes few assumptions about an adversary's goals and capabilities. This approach has a number of limitations [10]. Differential privacy is compositional—that is, running a differentially private mechanism twice also satisfies differential privacy, but at increased privacy cost. The compositionality of differential privacy separates it from a number of other privacy notions, including de-identification and k-anonymity. In both of those cases, two separate releases of data may individually satisfy the desired property, but may violate the property when taken together. Two differentially private releases of data, in contrast, may result in increased privacy cost, but will always satisfy differential privacy for some value of ε [2]. Differential privacy is designed to provide good utility for statistics about large populations in the data. Queries with low population size, by definition, pose an inherent privacy risk to individuals; differential privacy requires poor utility for their results in order to protect privacy [1]. That makes sense because, DP is often accompanied by significant loss in utility. While some of this loss is inherent, some may be incurred due to being overly cautious with what is and is not considered private. For example, suppose one wishes to learn how to diagnose patients based on a textual description of the patient. Some parts of this data may indeed be privacy sensitive: for example, characteristics of the patient and their symptoms. However, aspects such as grammar and syntax are basic parts of a language and are not privacy sensitive [10].

In the database realm, there are two types of DP. Unbounded differential privacy and bounded. In unbounded differential privacy, neighboring databases are formed by adding or removing a tuple in the database [2]. Bounded differential privacy. In bounded differential privacy, neighboring databases are formed by changing the value of exactly one tuple [2].

Regarding privacy, a primary goal of most database systems is to abstract away execution details, so that analysts may focus on the semantics of the queries they write instead of worrying about how they will be executed [2]. It is important to note that when working in a database that contains personally identifiable information or PII, nothing about an individual should be learnable from the database that cannot be learned without access to the database [19].

The main disadvantage of ensuring differential privacy is that it typically requires more noise infusion than traditional techniques. This is a consequence of the fact that traditional techniques only need to prevent linkage, while differential privacy prevents linkage through reconstruction. One might expect that in the discussion on how and when differential privacy should be applied, level-headed experts convene to weigh such pros and cons and find a consensus [3].

D. PPDM, PPGP & PPDP

Differential privacy practitioners understand the kinds of problems that DP can solve, such as releasing aggregate statistics on sensitive data, providing internal access to run queries on sensitive data to (semi-)trusted analysts, building and deploying ML models trained on sensitive data, and safely collecting telemetry data, but the developers, policy-makers, and business leaders who need to make decisions about what to do often have a hard time understanding the differences in assumptions and threat models be- tween these classes of use cases [10]. To protect from information leakage, privacy preservation methods have been developed to protect owner's exposure, by modifying the original data [9], [10]. However, transforming the data may also reduce its utility, resulting in inaccurate or even infeasible extraction of knowledge through data mining. This is the paradigm known as Privacy-Preserving Data Mining (PPDM). PPDM methodologies are designed to guarantee a certain level of privacy, while maximizing the utility of the data, such that data mining can still be performed on the transformed data efficiently. PPDM encompasses all techniques that can be used to extract knowledge from data while preserving privacy [9].

It is also obvious that statistics in the IT profession are increasingly useful. Statistics often involve the use of graphs and other visualization methods to convey meaning. In turn, PPPG's are also most noteworthy regarding privacy. PPGP, or privacy preserving graph publishing approaches can be broadly classified into five categories, namely graph modification techniques, graph generalization/clustering techniques, privacy aware graph computation techniques, differential privacy-based graph anonymity techniques, and hybrid anonymization techniques [4]. Regarding PPGP, a user's data anonymization is still irrefutably complex, and it requires significant improvements in existing approaches as well as devising new practical approaches with regard to better utility and privacy preservation [4]. Furthermore, Private statistical estimation problems are arguably one of the most prominent applications of data-adaptive DP algorithms. These tasks (e.g., private mean estimation) are, in general, impossible for worst-case datasets. To see why, observe that the mean of a dataset is arbitrarily sensitive to the addition or removal of a single extreme outlier. Even worse, the worst-case sensitivity is large for every dataset, even otherwise "well-behaved" datasets [10]. Differential privacy would prevent this, since outliers—and qualitative research in general—are by definition privacy-sensitive [3].

Regarding PPDP, or privacy preserving data publishing and overall security in general; data publishing privacy is achieved with privacy models that sanitize data. However, due to the access to other publicly available sources, adversaries can try to de-anonymize or to infer sensitive information [15], [16]. As the amount of published data continues to grow in both quantity and complexity, modelling background knowledge of adversaries presents several difficulties [17], such as the identification of what data can be used to de-anonymize and the amount of public data sources that can be linked together. This calls for the development of more evolved and realistic models of background knowledge available to adversaries, that can urge research on privacy mechanisms effective against these over- hauled adversaries [9].

IV. SQL

In the college curriculum and as a basic tool for database, it would be hard to dispute the practicality and usefulness of SQL. Although many forms and flavors of SQL exist, most share the same common query attribute and perform similar functions. While one institution may focus on Oracle, another may utilize Microsoft SQL, and yet another may choose the open-source nature of MySQL. In the area of DP, research exists that has integrated DP into distinct new forms of SQL.

The theory of differential privacy is being translated into practical systems for its deployment at an accelerating rate, and this process has uncovered a number of interesting challenges specific to the practical implementation of differentially private mechanisms [2]. Specifically, in alphabetical order, the following noteworthy implementations were uncovered during the research for this paper: Chorus, Flex, GoogleDP, PrivateSQL, and PINQ.

Chorus is a framework for building systems for differential privacy and requires more setup before deployment than systems like GoogleDP. However, Chorus works with any SQL database and supports a larger class of SQL-like queries than PINQ and GoolgeDP—in particular, queries that involve joins. Flex handles one query at a time and has a more efficient implementation of the Flex mechanism than the standard smooth sensitivity algorithm [2].

The Flex system (Johnson et al., 2018) uses an efficiently computed upper bound on local sensitivity, called elastic sensitivity, to bound the sensitivities of queries with general joins. The definition of elastic sensitivity assumes that all database records must be protected. In practice, databases often contain a mixture of sensitive and non-sensitive data. This fact can be used to tighten our bound on local sensitivity for queries joining on non- sensitive tables [1]. Elastic sensitivity does not require modifications to the data, and it can be easily applied as a post-processing step to an existing join query [2]. An open-source tool for computing elastic sensitivity of SQL queries was uncovered [23]. We use elastic sensitivity to build FLEX, a system for enforcing differential privacy for SQL queries. We evaluated FLEX on a wide variety of queries, demonstrating that FLEX can support real-world queries and provides high utility on a majority of queries with large population sizes [1]. Elastic sensitivity does not support non-equijoins, and adding support for these is not straightforward. Elastic sensitivity can also fail when requisite max-frequency metrics are not available due to the query structure [1].

GoogleDP (Wilson et al., 2020) handles a special type of privacy policy for multi-relational databases with constraints: user-level DP. The approach is like the truncation used to handle joins in PrivateSQL. If a single user may contribute k records to the database, then the stability of a base table is actually k—not 1. GoogleDP assumes that a single user may contribute many records and allows the analyst to specify a threshold k to bound the contribution of each user. The system applies a truncation rewrite to base tables that enforces the specified bound. The current implementation handles many queries, but no correlated subqueries. Hence, it works well for a simpler schema that has a well-defined "user" relation and simple SQL queries [2].

PrivateSQL supports complex privacy policies where neighboring databases differ more than one row due to the constraints between the primary private table and the secondary private tables. The PrivateSQL system (§ 7.5) uses a richer notion of neighboring databases based on database-specific privacy policies to specify privacy notions at multiple resolutions [2]. PrivateSQL is designed to meet three central goals: Workloads: The system should answer a workload of queries with bounded privacy loss. Complex Queries: Each query in the workload can be a complex SQL expression over multiple relations. Multi-resolution Privacy: The system should allow the data owner to specify which entities in the database re- quire protection [25].

Privacy Integrated Queries (PINQ) proposed by McSherry, 2009 is a platform that answers SQL-like queries on databases with a differential privacy guarantee. This platform is built on top of LINQ declarative query language. The techniques are generalizable to any SQL-like queries. For each query received, the platform automatically analyzes the query and then perturbs the query answer with the right amount of noise [2]. When to use PINQ. In PINQ, only a restricted form of JOIN is considered. This form of joins requires that each input data set is first grouped by its join keys, and the list of groups are then joined using their group keys. The result is a compact representation of the output of the original JOIN, as each pair of groups could in principle be expanded to their full Cartesian product. This type of join has bounded stability, as each input record participates in at most one pair of groups, and as with GROUP BY the stability constant is at most 2. However, this restriction limits each join key result in a single record no matter how large the group is. Hence, you cannot extract more information privately. It is still better than leaving the stability unbounded and this join is useful to link unique identifiers between data sets [2].

V. Hands on learning

There is an existing lesson on DP available at:

https://cloud.google.com/bigquery/docs/differential-privacy

This is useful if a risk of re-identification exists in your data. These will reduce the possibility that someone could make an inference about your data or groups of people, and they can prevent someone from learning something about an individual. [1] The fact that this is a hands-on lesson is extremely beneficial to learners. Google is a leader in the aspect of differential privacy. Google first deployed their world-class differential privacy anonymization technology in Chrome nearly seven years ago and are continually expanding its use across our products including Google Maps and the Assistant. And as the world combats COVID-19, last year we published our COVID-19 Community Mobility Reports, which uses differential privacy to help public health officials, economists and policymakers globally as they make critical decisions for their communities while ensuring no personally identifiable information is made available at any point [26].

VI. Descriptions of project work and methodology of research explorations

Our task of finding relevant journal articles, websites, publications, and syllabi included web searches, and utilizing available college curriculum course descriptions. Our research exploration was limited in technique but concise in its purpose of finding relevant information in the topics of curriculum, privacy concepts, and SQL based relational database management systems. In our exploration, we found that differential privacy is a promising approach to formalizing privacy—that is, for writing down what privacy means as a mathematical equation [2]. Through our research, we found that most of the work in differential privacy for databases has assumed that the data is fixed and does not change over time. In practice, however, databases do change over time—often continuously. This presents both challenges and opportunities for differential privacy [2].

VII. Research findings

In the realm of relational database classes, or curriculum, the survey of available research in the form of journal articles, publications, and websites, did not yield significant suitable hands-on learning opportunities for practical applications of DP. I would propose that conceptually, the research in this paper is suitable as a foundation of knowledge for DP for lecture presentation, with the hands-on lesson after introducing the concepts. Many of the concepts are very difficult to understand without a background in mathematics and programming. DP has seen wide deployment across industry and government organizations as the gold standard of privacy- preserving data analysis, but its mathematical precision makes its privacy guarantees difficult to understand. Simply describing it as "the gold standard" instills confidence but does not provide information about the nature of the privacy guarantees. On the other hand, describing it as "a bound on the worst-case ratio of the probability of a particular output of a randomized algorithm across two neighboring databases" is equally meaningless to those unfamiliar with the definition or without a mathematical background [10]. Hence, we must find a practical way to introduce these difficult concepts in an undergraduate database class.

VIII. Future scope of work

In the future for our undergraduate database classes, we will need to explore furthermore. We will need to provide deeper insights on the privacy problems in future computing paradigm that will be helpful in devising more secure anonymization methods [4]. In the past several years, driven by evolving functional and non-functional needs of an organization, alternatives to the classic relational model have emerged. Future employees need to be aware of the critical role of data privacy in every organization's analysis of big data, as well as the consequences that may ensue if efforts are not reasonably made to protect confidentiality [7]. As far as the future of big data, we should examine illustrative samples of these popular alternatives known as non-relational or NoSQL models [6]. Estimates are that more than 90% of the world's data is not structured (i.e., not in classical relational

databases amenable to SQL queries). What type of new actionable insights are facilitated by the processing of semi-structured (e.g., csv, JSON) and unstructured (e.g., text, images, audio) data [6]?

I would also propose that a privacy-based security class at the senior level or 300-400 level for most institutions; would also provide a great opportunity to integrate hands-on activities with DP, as well as other PET's. Students who already have a formal knowledge base in SQL would benefit more from the advanced concepts of DP in a hands-on learning situation as described in section V.

IX. Overall Summary and Conclusion

A unique aspect of my work is the consolidation of PET principals, DP techniques and privacy enhanced iterations of SQL. To my knowledge, this paper is one of the first of its kind to serve as a foundational knowledge guide for the introduction of DP into the undergraduate curriculum in the post IS2020 era. A paper on integrating privacy alone in undergraduate curriculum was found, however it lacked specificity to DP; and was speaking more in general curriculum aspects [27]. A few instances were found that were dated to 2016; but they lacked the specific examples of SQL that would be integral to a hands-on database class. Topics for lecture on practical, applicable information regarding DBMS's that will provide support for DP, as well as a hands-on lesson using DP provide a foundation of knowledge for DP.

References

[1] N. Johnson, J. P. Near, and D. Song, "Towards practical differential privacy for SQL queries," Proceedings of the VLDB Endowment, vol. 11, no. 5, pp. 526–539, Jan. 2018. [Online]. Available: https://doi.org/10.1145/3187009.3177733

[2] J. P. Near and X. He, "Differential Privacy for Databases," Foundations and Trends® in Databases, vol. 11, no. 2, pp. 109–225, 2021. [Online]. Available: https://doi.org/10.1561/1900000066

[3] D. L. Oberski and F. Kreuter, "Differential Privacy and Social Science: An Urgent Puzzle," vol. 2, no. 1, Jan. 2020. [Online]. Available: https://doi.org/10.1162/99608f92.63a22079

[4] A. Majeed and S. Lee, "Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey," IEEE Access, pp. 1–1, 2020. [Online]. Available: https://doi.org/10.1109/access.2020.3045700

[5] P. Leidig, R. Ferguson, and J. Reynolds, "Invited Paper: IS2010: A Retrospective Review and Recommendation," Journal of Information Systems Education, vol. 30, no. 4, pp. 298–302, 2019. [Online]. Available: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1356&context=jise

[6] "IS2020 A Competency Model for Undergraduate Programs in Information Systems." [Online]. Available: https://is2020.hosting2.acm.org/wp-content/uploads/2021/01/IS-2020-Final-Draft-Report.pdf

[7] D. Schwieger and C. Ladwig, "Protecting Privacy in Big Data: A Layered Approach for Curriculum Integration," Information Systems Education Journal (ISEDJ), vol. 14, no. 3, 2016. [Online]. Available: https://files.eric.ed.gov/fulltext/EJ1136177.pdf

[8] "Differential Privacy for Databases," DPforDB.github.io. [Online]. Available: https://dpfordb.github.io/ (accessed Jun. 03, 2023).

[9] R. Mendes and J. P. Vilela, "Privacy-Preserving Data Mining: Methods, Metrics, and Applications," IEEE Access, vol. 5, pp. 10562–10582, 2017. [Online]. Available: https://doi.org/10.1109/access.2017.2706947

[10] R. Cummings et al., "Challenges towards the Next Frontier in Privacy," arXiv.org, Apr. 14, 2023. [Online]. Available: https://arxiv.org/abs/2304.06929 (accessed Jun. 03, 2023).

[11] "Use differential privacy | BigQuery," Google Cloud. [Online]. Available: https://cloud.google.com/bigquery/docs/differential-privacy (accessed Jun. 03, 2023).

[12] "EMERGING PRIVACY ENHANCING TECHNOLOGIES CURRENT REGULATORY AND POLICY APPROACHES." [Online]. Available: [Privacy Enhancing Technologies Document](https://media.licdn.com/dms/document/media/C561FAQGCcpRsnWqBwQ/feedshare-document-pdf analyzed/0/1678738005167?e=1686787200&v=beta&t=tS9UI3PYFP15sVDRxDyXkS9wq8RhN_MYCXuRyhfsGp8)

[13] H. Topi et al., "IS 2010: Curriculum Guidelines for Undergraduate Degree Programs in Information Systems," Communications of the Association for Information Systems, vol. 26, 2010. [Online]. Available: https://doi.org/10.17705/1cais.02618

[14] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI Mag., vol. 17, no. 3, pp. 37–54, 1996.

[15] L. Sweeney, "K-anonymity: A model for protecting privacy," Int. J. Uncertainty, Fuzziness Knowl.-Based Syst., vol. 10, no. 5, pp. 557–570, 2002.

[16] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," in Proc. 30th IEEE Symp. Secur. Privacy, May 2009, pp. 173–187.

[17] B. Zhou, J. Pei, and W. S. Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data," ACM SIGKDD Explorations Newslett., vol. 10, no. 2, pp. 12–22, 2008.

[18] C. Dwork, "Differential Privacy," vol. 4052, 2006. [Online]. Available: https://www.microsoft.com/en-us/research/publication/differential-privacy/

[19] T. Dalenius, "Towards a methodology for statistical disclosure control," Statistik Tidskrift, 15, pp. 429–222, 1977.

[20] C. Dwork, "Differential privacy: A survey of results," in Proc. Int. Conf. Theory Appl. Models Comput., Changsha, China: Springer, 2008, pp. 1–19.

[21] X. Ding, W. Yang, K.-K. Raymond Choo, X. Wang, and H. Jin, "Privacy preserving similarity joins using MapReduce," Inf. Sci., vol. 493, pp. 20–33, Aug. 2019.

[22] F. Mcsherry, "Privacy Integrated Queries: An Extensible Platform for Privacy-Preserving Data Analysis," https://doi.org/10.1145/1810891

[23] "Overview," GitHub, Dec. 29, 2021. [Online]. Available: https://github.com/uber-archive/sql-differential-privacy

[24] WilliamDAssafMSFT, "Create Check Constraints - SQL Server," learn.microsoft.com. [Online]. Available: https://learn.microsoft.com/en-us/sql/relational-databases/tables/create-check-constraints?view=sql-server-ver16

[25] I. Kotsogiannis et al., "PrivateSQL," Proceedings of the VLDB Endowment, vol. 12, no. 11, pp. 1371–1384, Jul. 2019. [Online]. Available: https://doi.org/10.14778/3342263.3342274

[26] "How we're helping developers with differential privacy." https://developers.googleblog.com/2021/01/how-were-helping-developers-with-differential-privacy.html

[27] J. Vaidya, B. Shafiq, D. Lorenzi, and N. Badar, "Incorporating Privacy into the Undergraduate Curriculum," Oct. 2013, doi: https://doi.org/10.1145/2528908.2528918.