# Privacy and Disclosure Methods and Designs

## Future Directions for Social and Behavioral Science Methodologies in the Next Decade

Ruobin Gong

Department of Statistics
Rutgers University

Sept 25-26, 2024 · Washington, DC

National Academies of Sciences, Engineering, and Medicine
Division of Behavioral and Social Sciences and Education, Committee on National Statistics

# Formal Privacy in Statistical Disclosure Control

**Statistical Disclosure Control** (SDC) aims to limit particular measures of disclosure risk, e.g.:

- Reidentification: isolation of uniques [Swe02] ; model-based reidentification [FS98; SS08] ;
- Reconstruction: e.g. linkage attacks [DFT03; Win04] .

**Formal privacy** in SDC saw major development over the last decade or so.

- Principles [BG21; SM+24] :
  - Provability;
  - Composition;
  - Post-processing invariance;
  - Transparency (algorithmic ✓ ; statistical ? [Gon22] ).
- Differential privacy: a class of cryptographic standards. Crudely speaking [BGM24a; BGM24b] :

For all potential data universes $\mathcal{D} \in \mathscr{D}$ and protection objects $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$,

$$d_{\mathrm{P}_\Gamma}\left[\mathrm{P}_{\boldsymbol{x}}(T \in \cdot), \mathrm{P}_{\boldsymbol{x}'}(T \in \cdot)\right] \leq \epsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}').$$

# Formal Privacy in Statistical Disclosure Control

**Statistical Disclosure Control** (SDC) aims to limit particular measures of disclosure risk, e.g.:

- Reidentification: isolation of uniques [Swe02] ; model-based reidentification [FS98; SS08] ;
- Reconstruction: e.g. linkage attacks [DFT03; Win04] .

**Formal privacy** in SDC saw major development over the last decade or so.

- Principles [BG21; SM+24] :
  - Provability;
  - Composition;
  - Post-processing invariance;
  - Transparency (algorithmic ✓; statistical ? [Gon22] ).
- Differential privacy: a class of cryptographic standards. Crudely speaking [BGM24a; BGM24b] :

For all potential data universes $\mathcal{D} \in \mathscr{D}$ and protection objects $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}$,

$$d_{\mathrm{Pr}}\left[\mathsf{P}_{\boldsymbol{x}}(T \in \cdot), \mathsf{P}_{\boldsymbol{x}'}(T \in \cdot)\right] \leq \epsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}').$$

# Formal Privacy in Statistical Disclosure Control

The evolution of the differential privacy definition: catering to a variety of application domains

- From 2006: $\epsilon$-indistinguishability [Dwo+06b] ...
- ...to today:
  - Variations in output metric: $(\epsilon, \delta)$-approximate DP [Dwo+06a] Rényi DP [Mir17] concentrated DP [BS16] $f$-divergence privacy [BD14; BO13] $f$-DP (including Gaussian DP) [DRS22];
  - Variations in input metric: $(\mathcal{R}, \epsilon)$-generic DP [KM11] edge vs node privacy [Hay+09; MM10] $d$-metric DP [Cha+13] Blowfish privacy [HMD14] element level DP [ADJ22] distributional privacy [ZLW09] event-level vs user-level DP [Dwo+10];
  - Variations in the protection domain: privacy under invariants [Ash+19; GM20; GGY22; Dha+23] conditioned or empirical DP [ASV13; CH16] personalized DP [ESS15; JYC15] individual DP [Sor+17; FZ22] bootstrap DP [OC19] stratified DP [Bun+22] per-record DP [See+23] per-instance DP [Wan18; RW21];
  - Variations in applicable data structures: DP for network data [Hay+09] for geospatial data [And+13] Pufferfish DP [KM14] noiseless privacy [Bha+11] privacy under partial knowledge [SRS22] privacy amplification [BKN10; BBG20; Bun+22].

See generally [DP20].

# Formally Private Disclosure Mechanism Design: State of Research

In terms of modes of data access [Kar16; Hot+22] :

- Direct access (e.g. RDCs)
- Query access
- Dissemination access

In terms of the disclosure approach [SS23] :

- Design-based
- Adjustment-based

. . .

We know more about how to carry out a particular statistical estimation task (and publish the result) under an explicit formal privacy definition.

- e.g. point estimation [Smi11] , hypothesis testing [AS18; Can+19; Bar+19; BJ22] , confidence intervals [Du+20; KV18; WKL19; Dre+22; FWS22] , linear regression [She17; BS19; AV22] . . .
- But see [Bar+24] (“*DP methods are feasible for simple, univariate statistics but struggle to produce accurate regression estimates and confidence intervals*”)

We know less about how to design formally private mechanisms for the public dissemination of “raw data” that are well-suited for downstream use.

# Formally Private Disclosure Mechanism Design: State of Research

In terms of modes of data access [Kar16; Hot+22] :

- Direct access (e.g. RDCs)
- Query access (more)
- Dissemination access (less)

In terms of the disclosure approach [SS23] :

- Design-based (more)
- Adjustment-based (less)

. . .

We know more about how to carry out a particular statistical estimation task (and publish the result) under an explicit formal privacy definition.

- e.g. point estimation [Smi11] , hypothesis testing [AS18; Can+19; Bar+19; BJ22] , confidence intervals [Du+20; KV18; WKL19; Dre+22; FWS22] , linear regression [She17; BS19; AV22] . . .
- But see [Bar+24] ("*DP methods are feasible for simple, univariate statistics but struggle to produce accurate regression estimates and confidence intervals*")

We know less about how to design formally private mechanisms for the public dissemination of "raw data" that are well-suited for downstream use.

# Formally Private Disclosure Mechanism: Adoption in National Statistics

- Large-scale tabular datasets with consistency constraints
  - TopDown algorithm: 2020 U.S. Decennial Census P.L. and DHC files [Abo+22]
  - SafeTab: Detailed DHC files [Tum22]
- Differentially private synthetic data
  - NIST PSCR Differential Privacy Synthetic Data Challenge: see generally [BS21]
  - Israel's National Registry of Live Births [HC24]
- Synthetic data + differentially private verification or validation servers: toward tiered access
  - *Office of Personnel Management personnel records [Bar+18]
  - (under development) IRS Statistics of Income Division + Urban Institute [Bar+24]
- ? Large-scale longitudinal surveys, panel studies, complex surveys
  - (not yet) American Community Survey (ACS) [Uni22] ("*Our current assessment is that the science does not yet exist to comprehensively implement a formally private solution for the ACS*")
  - (not yet) Survey of Income and Program Participation (SIPP) [SM+24] ("*[T]he statistical tools have not yet advanced to the point that they could create a differentially private synthetic dataset for the size and complexity of SIPP*")

# Formally Private Disclosure Mechanism: Adoption in National Statistics

- Large-scale tabular datasets with consistency constraints
  - TopDown algorithm: 2020 U.S. Decennial Census P.L. and DHC files [Abo+22]
  - SafeTab: Detailed DHC files [Tum22]
- Differentially private synthetic data
  - NIST PSCR Differential Privacy Synthetic Data Challenge: see generally [BS21]
  - Israel's National Registry of Live Births [HC24]
- Synthetic data + differentially private verification or validation servers: toward tiered access
  - *Office of Personnel Management personnel records [Bar+18]
  - (under development) IRS Statistics of Income Division + Urban Institute [Bar+24]

? Large-scale longitudinal surveys, panel studies, complex surveys

- (not yet) American Community Survey (ACS) [Tum21] ("*Our current assessment is that the science does not yet exist to comprehensively implement a formally private solution for the ACS*")
- (not yet) Survey of Income and Program Participation (SIPP) [SM+24] ("*[T]he statistical tools have not yet advanced to the point that they could create a differentially private synthetic dataset for the size and complexity of SIPP*")

# Formally Private Disclosure Mechanism: Adoption in National Statistics

- Large-scale tabular datasets with consistency constraints
  - TopDown algorithm: 2020 U.S. Decennial Census P.L. and DHC files [Abo+22]
  - SafeTab: Detailed DHC files [Tum22]
- Differentially private synthetic data
  - NIST PSCR Differential Privacy Synthetic Data Challenge: see generally [BS21]
  - Israel's National Registry of Live Births [HC24]
- Synthetic data + differentially private verification or validation servers: toward tiered access
  - *Office of Personnel Management personnel records [Bar+18]
  - (under development) IRS Statistics of Income Division + Urban Institute [Bar+24]
- ? Large-scale longitudinal surveys, panel studies, complex surveys
  - (not yet) American Community Survey (ACS) [Uni22] ("*Our current assessment is that the science does not yet exist to comprehensively implement a formally private solution for the ACS*")
  - (not yet) Survey of Income and Program Participation (SIPP) [SM+24] ("*[T]he statistical tools have not yet advanced to the point that they could create a differentially private synthetic dataset for the size and complexity of SIPP*")

# Research Questions for the Near Future

1. Formally private SDL methods that are suited to the reality of official statistics
   - "Minimally invasive": compatibility with existing protocols and procedures [Das+22]
   - Tailored to the modes of data access and disclosure approaches
   - Inter-agency data sharing [Adv22] , data blending [SM+24] , downstream use [Ste+22] , and misuse (!)
2. A data processing "pipeline" view of formal privacy
   - Accounting for sampling, measurement, editing, imputation, nonresponse and other model-based or model-assisted adjustments
   - e.g. "privacy for free" [WFS15] ; amplification [BKN10; BBG20; Bun+22]
3. Privacy-usability tradeoff assessments that are . . .
   - . . .not under-inclusive: complexities of disclosure risk [Ken+21] , complexities of usability criteria
   - . . .not over-inclusive: resource constraints, practicality [Elk22]
4. Workable integration of rigid formal privacy standards with:
   - Sectoral (e.g. HIPAA, FERPA) and omnibus (GDPR-like) legal requirements
   - Articulated privacy and confidentiality directives (Five Safes [DRW16] , contextual integrity [Nis04] )
   - Intuitive expectations from data contributors and data users

# Research Questions for the Near Future

1. Formally private SDL methods that are suited to the reality of official statistics
   - "Minimally invasive": compatibility with existing protocols and procedures [Das+22]
   - Tailored to the modes of data access and disclosure approaches
   - Inter-agency data sharing [Adv22] , data blending [SM+24] , downstream use [Ste+22] , and misuse (!)
2. A data processing "pipeline" view of formal privacy
   - Accounting for sampling, measurement, editing, imputation, nonresponse and other model-based or model-assisted adjustments
   - e.g. "privacy for free" [WFS15] ; amplification [BKN10; BBG20; Bun+22]
3. Privacy-usability tradeoff assessments that are . . .
   - . . .not under-inclusive: complexities of disclosure risk [Sen+21] , complexities of usability criteria
   - . . .not over-inclusive: resource constraints, practicality [Elh22]
4. Workable integration of rigid formal privacy standards with:
   - Sectoral (e.g. HIPAA, FERPA) and omnibus (GDPR-like) legal requirements
   - Articulated privacy and confidentiality directives (Five Safes [DRW16] , contextual integrity [Nis04] )
   - Intuitive expectations from data contributors and data users

# Research Questions for the Near Future

1. Formally private SDL methods that are suited to the reality of official statistics
   - "Minimally invasive": compatibility with existing protocols and procedures [Das+22]
   - Tailored to the modes of data access and disclosure approaches
   - Inter-agency data sharing [Adv22] , data blending [SM+24] , downstream use [Ste+22] , and misuse (!)
2. A data processing "pipeline" view of formal privacy
   - Accounting for sampling, measurement, editing, imputation, nonresponse and other model-based or model-assisted adjustments
   - e.g. "privacy for free" [WFS15] ; amplification [BKN10; BBG20; Bun+22]
3. Privacy-usability tradeoff assessments that are . . .
   - . . .not under-inclusive: complexities of disclosure risk [Ken+21] , complexities of usability criteria
   - . . .not over-inclusive: resource constraints, practicality [Elt22]
4. Workable integration of rigid formal privacy standards with:
   - Sectoral (e.g. HIPAA, FERPA) and omnibus (GDPR-like) legal requirements
   - Articulated privacy and confidentiality directives (Five Safes [DRW16] , contextual integrity [Nis04] )
   - Intuitive expectations from data contributors and data users

# Research Questions for the Near Future

1. Formally private SDL methods that are suited to the reality of official statistics
   - "Minimally invasive": compatibility with existing protocols and procedures [Das+22]
   - Tailored to the modes of data access and disclosure approaches
   - Inter-agency data sharing [Adv22] , data blending [SM+24] , downstream use [Ste+22] , and misuse (!)
2. A data processing "pipeline" view of formal privacy
   - Accounting for sampling, measurement, editing, imputation, nonresponse and other model-based or model-assisted adjustments
   - e.g. "privacy for free" [WFS15] ; amplification [BKN10; BBG20; Bun+22]
3. Privacy-usability tradeoff assessments that are . . .
   - . . .not under-inclusive: complexities of disclosure risk [Ken+21] , complexities of usability criteria
   - . . .not over-inclusive: resource constraints, practicality [Elt22]
4. Workable integration of rigid formal privacy standards with:
   - Sectoral (e.g. HIPAA, FERPA) and omnibus (GDPR-like) legal requirements
   - Articulated privacy and confidentiality directives (Five Safes [DRW16] , contextual integrity [Nis04] )
   - Intuitive expectations from data contributors and data users

# References I


[Abo+22] John Abowd et al. “The 2020 Census Disclosure Avoidance System TopDown Algorithm”. In: *Harvard Data Science Review* Special Issue 2 (June 2022). DOI: 10.1162/99608f92.529e3cb9.

[ADJ22] Hilal Asi, John C. Duchi, and O. Javidbakht. “Element Level Differential Privacy: The Right Granularity of Privacy”. In: *AAAI Workshop on Privacy-Preserving Artificial Intelligence*. Association for the Advancement of Artificial Intelligence, 2022.

[Adv22] Advisory Committee on Data for Evidence Building. *Year 2 Report*. Tech. rep. 2022.

[And+13] Miguel E. Andrés, Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. “Geo-Indistinguishability: Differential Privacy for Location-Based Systems”. In: *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*. CCS ’13. New York, NY, USA: Association for Computing Machinery, Nov. 2013, pp. 901–914. ISBN: 978-1-4503-2477-9. DOI: 10.1145/2508859.2516735.

[AS18] Jordan Awan and Aleksandra Slavković. “Differentially private uniformly most powerful tests for binomial data”. In: *Advances in Neural Information Processing Systems* 31 (2018), pp. 4208–4218.

# References II


[Ash+19] Robert Ashmead, Daniel Kifer, Philip Leclerc, Ashwin Machanavajjhala, and William Sexton. *Effective privacy after adjusting for invariants with applications to the 2020 Census*. Tech. rep. 2019.

[ASV13] John M Abowd, Matthew J Schneider, and Lars Vilhuber. "Differential privacy applications to Bayesian and linear mixed model estimation". In: *Journal of Privacy and Confidentiality* 5.1 (2013).

[AV22] Daniel Alabi and Salil Vadhan. "Hypothesis Testing for Differentially Private Linear Regression". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 14196–14209.

[Bar+18] Andrés F Barrientos, Alexander Bolton, Tom Balmat, Jerome P Reiter, John M de Figueiredo, Ashwin Machanavajjhala, Yan Chen, Charley Kneifel, and Mark DeLong. "Providing access to confidential research data through synthesis and verification: An application to data on employees of the US federal government". In: (2018).

[Bar+19] Andrés F Barrientos, Jerome P Reiter, Ashwin Machanavajjhala, and Yan Chen. "Differentially private significance tests for regression coefficients". In: *Journal of Computational and Graphical Statistics* 28.2 (2019), pp. 440–453.

# References III


[Bar+24] Andrés F Barrientos, Aaron R Williams, Joshua Snoke, and Claire McKay Bowen. "A Feasibility Study of Differentially Private Summary Statistics and Regression Analyses with Evaluations on Administrative and Survey Data". In: *Journal of the American Statistical Association* 119.545 (2024), pp. 52–65.

[BBG20] Borja Balle, Gilles Barthe, and Marco Gaboardi. "Privacy Profiles and Amplification by Subsampling". In: *Journal of Privacy and Confidentiality* 10.1 (Jan. 2020). ISSN: 2575-8527. DOI: 10.29012/jpc.726.

[BD14] Rina Foygel Barber and John C. Duchi. *Privacy and Statistical Risk: Formalisms and Minimax Bounds*. http://arxiv.org/abs/1412.4451. Dec. 2014. DOI: 10.48550/arXiv.1412.4451. arXiv: 1412.4451 [cs, math, stat].

[BG21] Claire McKay Bowen and Simson Garfinkel. "Philosophy of differential privacy". In: *Notices of the American Mathematical Society* 68.10 (2021), pp. 1727–39.

[BGM24a] James Bailie, Ruobin Gong, and Xiao-Li Meng. "A Refreshment Stirred, Not Shaken (I): Building Blocks of Differential Privacy". In: *In preparation* (2024+).

[BGM24b] James Bailie, Ruobin Gong, and Xiao-Li Meng. "A Refreshment Stirred, Not Shaken (II): Can Swapping Be Differentially Private?" In: *In preparation* (2024+).

# References IV


[Bha+11] Raghav Bhaskar, Abhishek Bhowmick, Vipul Goyal, Srivatsan Laxman, and Abhradeep Thakurta. "Noiseless Database Privacy". In: *Advances in Cryptology – ASIACRYPT 2011*. Ed. by Dong Hoon Lee and Xiaoyun Wang. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2011, pp. 215–232. ISBN: 978-3-642-25385-0. DOI: 10.1007/978-3-642-25385-0_12.

[BJ22] Rina Foygel Barber and Lucas Janson. "Testing goodness-of-fit and conditional independence with approximate co-sufficient sampling". In: *The Annals of Statistics* 50.5 (2022), pp. 2514–2544.

[BKN10] Amos Beimel, Shiva Prasad Kasiviswanathan, and Kobbi Nissim. "Bounds on the Sample Complexity for Private Learning and Private Data Release". In: *Proceedings of the 7th Theory of Cryptography Conference, TCC 2010, Zurich, Switzerland*. Ed. by Daniele Micciancio. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, Feb. 2010, pp. 437–454. DOI: 10.1007/978-3-642-11799-2_26.

# References V


[BO13] Gilles Barthe and Federico Olmedo. "Beyond Differential Privacy: Composition Theorems and Relational Logic for f-Divergences between Probabilistic Programs". In: *Automata, Languages, and Programming*. Ed. by Fedor V. Fomin, Rūsiņš Freivalds, Marta Kwiatkowska, and David Peleg. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2013, pp. 49–60. ISBN: 978-3-642-39212-2. DOI: 10.1007/978-3-642-39212-2_8.

[BS16] Mark Bun and Thomas Steinke. "Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds". In: *Theory of Cryptography*. Ed. by Martin Hirt and Adam Smith. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2016, pp. 635–658. ISBN: 978-3-662-53641-4. DOI: 10.1007/978-3-662-53641-4_24.

[BS19] Garrett Bernstein and Daniel R Sheldon. "Differentially private Bayesian linear regression". In: *Advances in Neural Information Processing Systems* 32 (2019), pp. 525–535.

[BS21] Claire McKay Bowen and Joshua Snoke. "Comparative Study of Differentially Private Synthetic Data Algorithms from the NIST PSCR Differential Privacy Synthetic Data Challenge". In: *Journal of Privacy and Confidentiality* 11.1 (2021).

# References VI


[Bun+22] Mark Bun, Jörg Drechsler, Marco Gaboardi, Audra McMillan, and Jayshree Sarathy. "Controlling Privacy Loss in Sampling Schemes: An Analysis of Stratified and Cluster Sampling". In: *Foundations of Responsible Computing (FORC 2022)*. June 2022, p. 24.

[Can+19] Clément L Canonne, Gautam Kamath, Audra McMillan, Adam Smith, and Jonathan Ullman. "The structure of optimal private tests for simple hypotheses". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. 2019, pp. 310–321.

[CH16] Anne-Sophie Charest and Yiwei Hou. "On the meaning and limits of empirical differential privacy". In: *Journal of Privacy and Confidentiality* 7.3 (2016), pp. 53–66.

[Cha+13] Konstantinos Chatzikokolakis, Miguel E. Andrés, Nicolás Emilio Bordenabe, and Catuscia Palamidessi. "Broadening the Scope of Differential Privacy Using Metrics". In: *Privacy Enhancing Technologies*. Ed. by Emiliano De Cristofaro and Matthew Wright. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2013, pp. 82–102. DOI: 10.1007/978-3-642-39077-7_5.

[Das+22] Soumojit Das, Jorg Drechsler, Keith Merrill, and Shawn Merrill. "Imputation under differential privacy". In: *arXiv preprint arXiv:2206.15063* (2022).

# References VII


[DFT03] Josep Domingo-Ferrer and Vicenc Torra. “Disclosure risk assessment in statistical microdata protection via advanced record linkage”. In: *Statistics and Computing* 13 (2003), pp. 343–354.

[Dha+23] Prathamesh Dharangutte, Jie Gao, Ruobin Gong, and Fang-Yi Yu. “Integer Subspace Differential Privacy”. In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-23)*. 2023.

[DP20] Damien Desfontaines and Balázs Pejó. “SoK: Differential Privacies”. In: *Proceedings on Privacy Enhancing Technologies* 2020.2 (2020), pp. 288–313.

[Dre+22] Jörg Drechsler, Ira Globus-Harris, Audra Mcmillan, Jayshree Sarathy, and Adam Smith. “Nonparametric differentially private confidence intervals for the median”. In: *Journal of Survey Statistics and Methodology* 10.3 (2022), pp. 804–829.

[DRS22] Jinshuo Dong, Aaron Roth, and Weijie J. Su. “Gaussian Differential Privacy”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 84.1 (2022), pp. 3–37. ISSN: 1467-9868. DOI: 10.1111/rssb.12454.

[DRW16] Tanvi Desai, Felix Ritchie, and Richard Welpton. *Five Safes: Designing Data Access for Research*. Working Paper 1601. Bristol: University of the West of England, 2016, p. 27.

# References VIII


[Du+20] Wenxin Du, Canyon Foot, Monica Moniot, Andrew Bray, and Adam Groce. "Differentially private confidence intervals". In: *arXiv preprint arXiv:2001.02285* (2020).

[Dwo+06a] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. "Our Data, Ourselves: Privacy Via Distributed Noise Generation". In: *Advances in Cryptology - EUROCRYPT 2006*. Ed. by Serge Vaudenay. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2006, pp. 486–503. ISBN: 978-3-540-34547-3. DOI: 10.1007/11761679_29.

[Dwo+06b] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. "Calibrating noise to sensitivity in private data analysis". In: *Theory of cryptography conference*. Springer. 2006, pp. 265–284.

[Dwo+10] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum. "Differential Privacy under Continual Observation". In: *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*. STOC '10. https://dl.acm.org/doi/10.1145/1806689.1806787. New York, NY, USA: Association for Computing Machinery, June 2010, pp. 715–724. ISBN: 978-1-4503-0050-6. DOI: 10.1145/1806689.1806787.

[Elt22] John L Eltinge. "Disclosure protection in the context of statistical agency operations: Data quality and related constraints". In: *Harvard Data Science Review. Published online June* 24 (2022).

# References IX


[ESS15] Hamid Ebadi, David Sands, and Gerardo Schneider. "Differential Privacy: Now It's Getting Personal". In: *ACM SIGPLAN Notices* 50.1 (Jan. 2015), pp. 69–81. ISSN: 0362-1340. DOI: 10.1145/2775051.2677005.

[FS98] Stephen E Fienberg and Russell J Steele. "Disclosure limitation using perturbation and related methods for categorical data". In: *Journal of Official Statistics* 14.4 (1998), p. 485.

[FWS22] Cecilia Ferrando, Shufan Wang, and Daniel Sheldon. "Parametric bootstrap for differentially private confidence intervals". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2022, pp. 1598–1618.

[FZ22] Vitaly Feldman and Tijana Zrnic. *Individual Privacy Accounting via a Rényi Filter*. http://arxiv.org/abs/2008.11193. Jan. 2022. arXiv: 2008.11193 [cs, stat].

[GGY22] Jie Gao, Ruobin Gong, and Fang-Yi Yu. "Subspace Differential Privacy". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. 4. June 2022, pp. 3986–3995. DOI: 10.1609/aaai.v36i4.20315.

# References X


[GM20] Ruobin Gong and Xiao-Li Meng. "Congenial differential privacy under mandated disclosure". In: *Proceedings of the 2020 ACM-IMS on Foundations of Data Science Conference*. FODS '20. 2020, pp. 59–70.

[Gon22] Ruobin Gong. "Transparent Privacy Is Principled Privacy". In: *Harvard Data Science Review* Special Issue 2 (June 2022). DOI: 10.1162/99608f92.b5d3faaa.

[Hay+09] Michael Hay, Chao Li, Gerome Miklau, and David Jensen. "Accurate Estimation of the Degree Distribution of Private Networks". In: *2009 Ninth IEEE International Conference on Data Mining*. Dec. 2009, pp. 169–178. DOI: 10.1109/ICDM.2009.11.

[HC24] Shlomi Hod and Ran Canetti. *Differentially Private Release of Israel's National Registry of Live Births*. 2024. arXiv: 2405.00267 [cs.CR]. URL: https://arxiv.org/abs/2405.00267.

[HMD14] Xi He, Ashwin Machanavajjhala, and Bolin Ding. "Blowfish privacy: Tuning privacy-utility trade-offs using policies". In: *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. 2014, pp. 1447–1458.

# References XI


[Hot+22] V Joseph Hotz, Christopher R Bollinger, Tatiana Komarova, Charles F Manski, Robert A Moffitt, Denis Nekipelov, Aaron Sojourner, and Bruce D Spencer. "Balancing data privacy and usability in the federal statistical system". In: *Proceedings of the National Academy of Sciences* 119.31 (2022), e2104906119.

[JYC15] Zach Jorgensen, Ting Yu, and Graham Cormode. "Conservative or Liberal? Personalized Differential Privacy". In: *2015 IEEE 31st International Conference on Data Engineering*. https://ieeexplore.ieee.org/document/7113353. Apr. 2015, pp. 1023–1034. DOI: 10.1109/ICDE.2015.7113353. (Visited on 09/30/2023).

[Kar16] Alan F Karr. "Data sharing and access". In: *Annual Review of Statistics and Its Application* 3.1 (2016), pp. 113–132.

[Ken+21] Christopher T Kenny, Shiro Kuriwaki, Cory McCartan, Evan TR Rosenman, Tyler Simko, and Kosuke Imai. "The use of differential privacy for census data and its impact on redistricting: The case of the 2020 US Census". In: *Science advances* 7.41 (2021), eabk3283.

[KM11] Daniel Kifer and Ashwin Machanavajjhala. "No Free Lunch in Data Privacy". In: *Proceedings of the 2011 International Conference on Management of Data - SIGMOD '11*. Athens, Greece: ACM Press, 2011, pp. 193–204. ISBN: 978-1-4503-0661-4. DOI: 10.1145/1989323.1989345.

# References XII


[KM14] Daniel Kifer and Ashwin Machanavajjhala. "Pufferfish: A framework for mathematical privacy definitions". In: *ACM Transactions on Database Systems (TODS)* 39.1 (2014), pp. 1–36.

[KV18] Vishesh Karwa and Salil Vadhan. "Finite Sample Differentially Private Confidence Intervals". In: *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2018.

[Mir17] Ilya Mironov. "Rényi Differential Privacy". In: *2017 IEEE 30th Computer Security Foundations Symposium (CSF)* (Aug. 2017), pp. 263–275. DOI: 10.1109/CSF.2017.11. eprint: 1702.07476. (Visited on 01/14/2020).

[MM10] Frank McSherry and Ratul Mahajan. "Differentially-Private Network Trace Analysis". In: *Proceedings of the ACM SIGCOMM 2010 Conference*. SIGCOMM '10. New York, NY, USA: Association for Computing Machinery, Aug. 2010, pp. 123–134. ISBN: 978-1-4503-0201-2. DOI: 10.1145/1851182.1851199.

[Nis04] Helen Nissenbaum. "Privacy as contextual integrity". In: *Wash. L. Rev.* 79 (2004), p. 119.

[OC19] Christine M O'Keefe and Anne-Sophie Charest. "Bootstrap differential privacy". In: *Transactions on Data Privacy* 12 (2019), pp. 1–28.

# References XIII


[RW21] Rachel Redberg and Yu-Xiang Wang. "Privately Publishable Per-Instance Privacy". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 17335–17346. (Visited on 03/29/2023).

[See+23] Jeremy Seeman, William Sexton, David Pujol, and Ashwin Machanavajjhala. "Per-Record Differential Privacy: Modeling Dependence between Individual Privacy Loss and Confidential Records". In: (2023+).

[She17] Or Sheffet. "Differentially private ordinary least squares". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 3105–3114.

[SM+24] Engineering National Academies of Sciences, Medicine, et al. *A Roadmap for Disclosure Avoidance in the Survey of Income and Program Participation*. 2024.

[Smi11] Adam Smith. "Privacy-preserving statistical estimation with optimal convergence rates". In: *Proceedings of the forty-third annual ACM symposium on Theory of computing*. 2011, pp. 813–822.

# References XIV


[Sor+17] Jordi Soria-Comas, Josep Domingo-Ferrer, David Sánchez, and David Megías. "Individual Differential Privacy: A Utility-Preserving Formulation of Differential Privacy Guarantees". In: *IEEE Transactions on Information Forensics and Security* 12.6 (June 2017), pp. 1418–1429. ISSN: 1556-6013, 1556-6021. DOI: 10.1109/TIFS.2017.2663337. (Visited on 03/29/2023).

[SRS22] Jeremy Seeman, Matthew Reimherr, and Aleksandra Slavkovic. *Formal Privacy for Partially Private Data*. http://arxiv.org/abs/2204.01102. May 2022. arXiv: 2204.01102 [cs, stat].

[SS08] Chris Skinner and Natalie Shlomo. "Assessing identification risk in survey microdata using log-linear models". In: *Journal of the American Statistical Association* 103.483 (2008), pp. 989–1001.

[SS23] Aleksandra Slavković and Jeremy Seeman. "Statistical data privacy: A song of privacy and utility". In: *Annual Review of Statistics and Its Application* 10.1 (2023), pp. 189–218.

[Ste+22] Ryan Steed, Terrance Liu, Zhiwei Steven Wu, and Alessandro Acquisti. "Policy impacts of statistical uncertainty and privacy". In: *Science* 377.6609 (2022), pp. 928–931.

[Swe02] Latanya Sweeney. "k-anonymity: A model for protecting privacy". In: *International journal of uncertainty, fuzziness and knowledge-based systems* 10.05 (2002), pp. 557–570.

# References XV


[Tum22] Tumult Labs. *SafeTab: DP Algorithms for 2020 Census Detailed DHC Race & Ethnicity*. Tech. rep. Mar. 2022.

[Uni22] United States Census Bureau. *Disclosure Avoidance Protections for the American Community Survey*. https://www.census.gov/newsroom/blogs/random-samplings/2022/12/disclosure-avoidance-protections-acs.html. 2022.

[Wan18] Yu-Xiang Wang. *Per-Instance Differential Privacy*. http://arxiv.org/abs/1707.07708. Nov. 2018. arXiv: 1707.07708 [cs, stat].

[WFS15] Yu-Xiang Wang, Stephen Fienberg, and Alex Smola. "Privacy for free: Posterior sampling and stochastic gradient monte carlo". In: *International Conference on Machine Learning*. PMLR. 2015, pp. 2493–2502.

[Win04] William E Winkler. "Re-identification methods for masked microdata". In: *International Workshop on Privacy in Statistical Databases*. Springer. 2004, pp. 216–230.

[WKL19] Yue Wang, Daniel Kifer, and Jaewoo Lee. "Differentially Private Confidence Intervals for Empirical Risk Minimization". In: *Journal of Privacy and Confidentiality* 9.1 (2019).

# References XVI

[ZLW09] Shuheng Zhou, Katrina Ligett, and Larry Wasserman. “Differential Privacy with Compression”. In: *Proceedings of the 2009 IEEE International Conference on Symposium on Information Theory - Volume 4*. ISIT’09. Coex, Seoul, Korea: IEEE Press, June 2009, pp. 2718–2722. ISBN: 978-1-4244-4312-3.