SoK: Considerations in Measuring Compliance with Privacy Regulations

Nathan Reitinger and Michelle L. Mazurek
University of Maryland

***Abstract*—Data privacy regulations, like the General Data Protection Regulation (GDRP) and the California Consumer Privacy Act (CCPA), provide transparency-focused privacy rights which are enforceable by certain types of individuals against certain types of data stewards. But how well are these regulations' mandates put into practice? To help answer this question, researchers have conducted measurement-style studies assessing legal compliance, but these efforts may not be as straightforward as they seem—regulations are complex, often ambiguous, and uniquely consequential, meaning that compliance is a difficult-to-answer question with ethical and legal ramifications. Here, we propose to develop a general framework for helping researchers identify and think through some key considerations when conducting this research.**

Introduction. Following on from the GDPR [1], U.S. states like California, Virginia, Colorado, Connecticut, and Utah have all enacted comprehensive, data-protection regulations [2]–[7]. These state-based laws grant similar rights as the GDPR, and pending legislation in at least four other states is set to do the same [2]. This is also true internationally, with 16 countries, so far, adopting GDPR-like schemes [8]. On top of this, sector-specific statutes in the U.S., like BIPA, COPPA, and HIPAA, also mandate related privacy-protective practices [9]–[11].

As these regulations proliferate, the research community increasingly seeks to measure their impact, including rates and types of compliance (e.g., [12]–[24]). These papers typically leverage large-scale analysis, highlight discrepancies in compliance, and provide guidance to data stewards who must comply with the regulation being considered.

However, compliance itself is a nontrivial question. Legal texts like statutes are famously complex, ambiguous, and leave room for interpretive debate [25]–[30], meaning that definitive statements about compliance can be difficult to produce, and, if not undertaken carefully, give regulated entities an "easy out" when disputing a claim of non-compliance. Compliance also implies real-world consequence (e.g., monetary penalties or reputational harm), meaning that ethics plays an important role: deeming specific entities non-compliant can have serious consequences, underscoring the importance of avoiding false positives—even the process of measuring compliance may cause undue anxiety [31]–[33] or require non-trivial effort from recipients.[1] As such, before conducting this type of research, it is important to understand and evaluate the pros, cons, and trade-offs of different measurement approaches.

Methods. We will start by creating a dataset of academic papers measuring legal compliance. We plan on using two data sources: (1) the ACM digital library [34]; and (2) collecting all papers in the past five years from specific, relevant conferences and then filtering using keywords. To search the ACM digital library, we will use a keyword search on the "short" names of privacy-protective regulations (e.g., GDPR, CCPA, VCDPA) together with the word "compliance" (e.g., <CCPA> <compliance>). In our preliminary testing, this search produced relevant papers which were outside the scope of typical US-based conferences (important for studying the GDPR), but relevant to our research interests. We will also collect all papers from the past five years from conferences such as: PETS, WWW, WPES, CHI, SOUPS, CCS, USENIX, S&P, and NDSS, and then filter them using the same short-statute + "compliance" search. All papers will be further filtered manually, based on titles and abstracts, assessing relevance broadly.

Compliance Framework. Next, we will inductively develop a codebook by analyzing, in detail, a subset of papers aligned with our research goals. We expect to generate codes relating to measurement methods, definitions of compliance, and ethical considerations. Potential examples include:

- How sender was deemed to comply with target law
- How recipient was deemed to comply with target law
- Use of statutory definitions, case law, or other approaches to define compliance
- Approach to identifying ground truth used to measure the recipient's response
- Was the research considered by an IRB or equivalent committee (possibly unnecessary or inapplicable)
- How and why were recipients informed or debriefed about the goals and scope of the research
- How did researchers consider effort and stress for recipients

Goals. Measuring legal compliance is complicated; our goal is to identify key challenges, trade-offs, and methodologies in this area. We do not intend to offer a specific per-regulation guide, or to suggest that there are any correct one-size-fits-all answers, but rather to help future researchers in planning their measurement studies effectively. We seek feedback and potential collaboration from experts in the field.

1. "Recipient" means the regulated entity. "Sender" means the individuals exercising a privacy right. "Compliance" means verification of a statute's mandates (e.g., fulfilling a "right to know" request).

References

[1] "Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)," *L 119, Official journal of the European Union*, pp. 1–88, 2016.

[2] IAPP. US state privacy legislation tracker. https://iapp.org/media/pdf/resource_center/State_Comp_Privacy_Law_Chart.pdf.

[3] "California Civil Code," *§§ 1798.100–1798.199.100*, 2018.

[4] "Virginia Consumer Data Protection Act (VCDPA)," *Virginia Code Annotated §§ 59.1-575–59.1-585*, 2021.

[5] "Colorado Privacy Act (CPA)," *Colorado Revised Statutes §§ 6-1-1301–6-1-1313*, 2021.

[6] "Connecticut Personal Data Privacy and Online Monitoring Act (CTDPA)," *Connecticut General Statutes §§ 42-515–42-526*, 2022.

[7] "Utah Consumer Privacy Act (UCPA)," *Utah Code Annotated §§ 13-61-401–13-61-404*, 2022.

[8] Dan Simmons, "17 countries with GDPR-like data privacy laws," https://insights.comforte.com/countries-with-gdpr-like-data-privacy-laws, EDPS, 2022.

[9] "Illinois Biometric Information Privacy Act (BIPA)," *740 Illinois Compiled Statutes §§ 14/1–14/99*, 2020.

[10] "Children's Online Privacy Protection Act (COPPA)," *15 U.S.C. §§ 6501–6506*, 2006.

[11] "Health Insurance Portability and Accountability Act of 1996 (HIPAA)," *Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered sections of 18, 26, 29, and 42 U.S.C.)*, 2012.

[12] Sean O'Connor, Ryan Nurwono, Aden Siebel, and Eleanor Birrell, "(Un)clear and (in)conspicuous: the right to opt-out of sale under CCPA." In *Proc. WPES*, 2021.

[13] Yevgeniy Dodis, Jonathan Katz, Adam Smith, and Shabsi Walfish, "Fighting the fog: evaluating the clarity of privacy disclosures in the age of CCPA." In *Proc. WPES*, 2021.

[14] Fuman Xie, Yanjun Zhang, Chuan Yan, Suwan Li, Lei Bu, Kai Chen, Zi Huang, and Guangdong Bai, "Scrutinizing privacy policy compliance of virtual personal assistant apps." In *Proc. ASE*, 2022.

[15] Natalija Vlajic, Marmara El Masri, Gianluigi M. Riva, Marguerite Barry and Derek Doran, "Online Tracking of kids and teens by means of invisible Images: COPPA vs. GDPR." In *Proc. MPS*, 2018.

[16] Shuang Liu, Baiyang Zhao, Renjie Guo, Guozhu Meng, Fan Zhang and Meishan Zhang, "Have you been properly notified? automatic compliance analysis of privacy policy text with GDPR Article 13." In *Proc. WWW*, 2021.

[17] Razieh Nokhbeh Zaeem, K. Suzanne Barber, "The effect of the GDPR on privacy policies: recent progress and future promise." In *Proc. CCS*, 2020.

[18] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez and Serge Egelman, "'Won't somebody think of the children?' Examining COPPA compliance at scale." In *Proc. PETS*, 2018.

[19] Tobias Urban, Dennis Tatang, Martin Degeling, Thorsten Holz and Norbert Pohlmann, "A study on subject data access in online advertising after the GDPR," *LNSC*, vol. 11737, pp. 61–79, 2019.

[20] Maggie Van Nortwick and Christo Wilson, "Setting the bar low: are websites complying with the minimum requirements of the CCPA?" In *Proc. PETS*, 2022.

[21] Tamjid Al Rahat, Minjun Long and Yuan Tian, "Is your policy compliant? A deep learning-based empirical study of privacy policies' compliance with GDPR." In *Proc. WPES*, 2022.

[22] Jacob Leon Kröger, Jens Lindemann and Dominik Herrmann, "How do app vendors respond to subject access requests? A longitudinal privacy study on iOS and Android Apps." In *Proc. ARES*, 2020.

[23] Yuxi Ling, Kailong Wang, Guangdong Bai, Haoyu Wang and Jin Song Dong, "Are they toeing the line? Diagnosing privacy compliance violations among browser extensions." In *Proc. ARES*, 2022.

[24] Mariano Di Martino, Pieter Robyns, Winnie Weyts, Peter Quax, Wim Lamotte and Ken Andries, "Personal information leakage by abusing the GDPR 'right of access'." In *Proc. SOUPS*, 2019.

[25] Aleecia M McDonald and Lorrie Faith Cranor, "The cost of reading privacy policies," *I/S: A journal of law and policy for the information society*, vol. 4, pp. 543–568, 2008.

[26] Jonathan A. Obar and Anne Oeldorf-Hirsch, "The biggest lie on the internet: ignoring the privacy policies and terms of service policies of social networking services," *Information, communication & society*, vol. 23, no. 1, pp. 128–147, 2020.

[27] Ryan Amos, Gunes Acar, Eli Lucherini, Mihir Kshirsagar, Arvind Narayanan and Jonathan Mayer, "Privacy policies over time: curation and analysis of a million-document dataset." In *Proc. WWW*, 2021.

[28] Michael Curtotti and Eric McCreath, "A Right to Access Implies A Right to Know: An Open Online Platform for Research on the Readability of Law," *Data organization and legal informatics*, vol. 1, pp. 1–56, 2013.

[29] Ward Farnsworth1, Dustin F. Guzior and Anup Malani, "Ambiguity about ambiguity: an empirical inquiry into legal interpretation," *Journal of legal analysis*, vol. 2, pp. 257–300, 2010.

[30] "Ambiguity and misunderstanding in the law," *Thomas Jefferson Law Review*, vol. 25, pp. 167–193, 2002.

[31] Jon Baines, "Data ethics concerns halt academic study into subject access requests," https://www.mishcon.com/news/data-ethics-concerns-halt-academic-study-into-subject-access-requests, Mishcon de Reya, 2021.

[32] Kirk Strauser, "Dealing With Princeton's Flawed Privacy Research," https://honeypot.net/post/dealing-with-princetons-flawed-privacy-research/, 2021.

[33] Jonathan Mayer, "Princeton-Radboud study on privacy law implementation," https://privacystudy.cs.princeton.edu, 2021.

[34] ACM Digital Library, "Advanced Search," https://dl.acm.org/search/advanced.