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 (GDPR) 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 [29]–[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.
References


