Understanding Organizational Dynamics of Federated Data Collection and Privacy Preservation

David Loshin	Brian Butler
PhD Student	Senior Associate Dean and Professor
University of Maryland -- College of Information Studies	Co-Director, Social Data Science Center (SoDa)
College Park, MD USA	University of Maryland -- College of Information Studies
dloshin@umd.edu	College Park, MD USA
	bsbutler@umd.edu

As organizations have joined the fight to slow the community spread of COVID-19, their leaders find that they are now not only in a position to collect and manage individuals' data at a scale to which they are not accustomed, they also are expected to interoperate in a loosely federated network to share information, comply with quickly defined and passed public health mandates such as contact tracing and reporting, collaborate with organizations in similar situations, all while providing assurance of compliance with data privacy protection laws with which they have little experience.

Organizations, such as healthcare providers, that have traditionally relied on centralized information management and complex regulatory systems have more easily adapted to the demands of contact tracing and its associated reporting. In a more tightly federated environment where organizations that are working with the data are familiar with regulations and norms, the value and cost structures facilitate compliance - if the hospital is already subject to governed data management and protection practices (e.g. HIPAA Privacy Rule), then it is relatively easier to layer in contact tracing and reporting.

However, for contact tracing to provide the optimal public health benefit, similar public health practices (monitoring COVID symptoms, requiring testing, reporting) must be implemented across a highly decentralized, loosely federated network of organizations with varying degrees of experience (such as universities, public school systems, private schools, restaurants, movie theaters, houses of worship). These organizations are expected to request information and collect information from its participants (staff and constituents alike) as well as report COVID infections to health authorities while observing data privacy of protected health information.

These types of organizations have widely varying infrastructure for either the processes or the technologies needed to facilitate compliance, imposing additional cost and operational constraints. If an organization has never had to institute health-related reporting processes and must do so under crisis conditions, the base assessments of cost value and risk will significantly affect both compliance and quality.

The organizational dynamics of federated data collection and privacy preservation approaches are also complicated by the different relationships that organizations have with their stakeholders. An organization can exercise some control over the behavior of its staff but is only positioned to *influence* behaviors of its customers or constituents. Even if the constituents agree to comply with directives

within the context of the organization, the organization has little control over the decisions constituents make when they are outside the context. For example, a university can impose rules about self-monitoring and logging of symptoms, COVID-19 testing, social distancing within the classroom, and wearing masks on campus, but it has little control over the individual student's decision to attend a party where these precautions are not observed. Other factors also impact organizational decision-making about contact tracing and reporting. For example, a restaurant bent on surviving the economic downturn might be more lenient in monitoring customer symptoms and reporting to ensure its revenue stream. This scenario, which is only going to increasingly replicate as more regions open their economies, poses challenges that are only partially addressed by data management and integrated contact tracing technologies, and raises questions about how systems and networks of organizations manage their decision-making and processes for privacy protection, such as:

- How is organizational decision-making about privacy and data collection impacted by the scope of its sphere of influence over individual actions?
- How do individual stakeholder motivations impact the organization's privacy and data collection practices and process?
- How do constituent/participant behaviors inform and influence the organizational decision-making about privacy and data collection ?
- How do organizations manage their privacy and data collection decisions in concert with other similar organizations?
- How are data protection and privacy norms modulated across networks of organizations interoperating in this loose and sometimes hierarchical federation?
- What roles do organizations play in shaping one another's the privacy and data collection behaviors?

Ultimately, how privacy policies and practices are implemented is inextricably linked to the ways that organizations collaborate, compete, make decisions, and change. Almost every organizations' ability to reopen and operate are now dependent on its ability to navigate the challenges of collecting and appropriately using health-related data. Understanding how diverse organizations make data collection, use, and sharing decisions in the context of others making decisions is essential for understanding how individual concerns about privacy will be balanced with organizational and societal needs. Supporting responsible uses of health data during times of crisis involves shaping organizational decisions, processes, and practices.