Privacy-Preserving Data Governance

Riley Wong

MARCH 2024

HARVARD Kennedy School
ASH CENTER
for Democratic Governance
and Innovation

How can privacy and cryptography tools enable data consent, models beyond ownership, and collective power?

Emerging models for collective data governance, including data trusts, data co-ops, and data coalitions, create opportunities for new forms and paradigms for agency over how data can be accessed, used, and managed. Progress in cryptographic techniques, including zero-knowledge proofs, multi-party computations, fully homomorphic encryption, verifiable credentials, and decentralized identity, enables new models focused on community privacy. How can we build up an ecosystem that integrates these frameworks, especially in service of underserved communities, such as sex workers, QTBIPOC (queer and trans Black, Indigenous, and people of color), labor unions, journalists, populations under authoritarian regimes, and more?

Emerging Technologies for Privacy and Cryptography

End-to-end encryption has made private and secure messaging possible. With encrypted messaging services like Signal, sex workers, community organizers, protesters, activists, journalists, labor unions, and many other vulnerable communities can access trusted communications, securely spread information, and transmit private data with a reduced risk of malicious interference, government surveillance, or other forms of commercial or invasive tracking.

In a similar way, how can cryptographic guarantees from these emerging technologies—such as zero-knowledge proofs, multi-party computations, and fully homomorphic encryption—enable new forms of privacy protections and create community infrastructures that promote collective agency and consent?

Zero-Knowledge Proofs

Zero-knowledge (ZK) proofs are cryptographic protocols that allow one party, the prover, to guarantee to another party, the verifier, that they know a certain piece of information without revealing that information itself. This makes it possible to prove the truth value of a piece of information without revealing the information itself. For example, you could prove membership of a specific group, verify a specific credential, or grant access rights while maintaining privacy of identity.

One example of how zero-knowledge proofs can benefit the sex worker ecosystem is by enabling privacy-preserving age verification for accessing sex worker platforms. States including North Carolina, Utah, Mississippi, Virginia, and Louisiana have recently passed state laws that require age verifications to access porn sites. Without ZK proofs, many existing age verification solutions may involve storing personal data, such as government IDs, credit card information, or face scans, in commercial databases. These may be distributed across multiple databases that may or may not be secure. With the privacy protections enabled by ZK proofs, one can simply provide a cryptographic proof to websites without disclosing any personal information besides the verification that one is at least 18 years of age.

ZK proofs may also benefit sex workers by facilitating a system where both clients and sex workers can provide a "proof of status" without revealing any other medical information or divulging ties between their medical results and personal identifiers. To do this now, one might have to show a test result that includes their legal name or other personal information. With a zero-knowledge system, it becomes possible to prove the status of any sexually transmitted infections while maintaining privacy of one's identity. This can promote safety, privacy, and anonymity while enabling all parties to make informed, consensual decisions around appropriate precautions or protections for collective sexual wellness.

Multi-Party Computation

Multi-party computation (MPC) is a cryptographic technique that allows multiple parties to jointly compute a function or perform a calculation over their inputs while maintaining the privacy of each input. No party has to reveal their private input to any other party in order to get the computational output.

MPCs can be a useful tool for matching algorithms, such as price matching or services matching. For example, a sex worker offers a list of services and wants to keep these offerings private unless a specific service is requested. At the same time, a client may be seeking a set of services without necessarily wanting to broadcast what these are unless they are offered. Using MPC, we can reveal only the services that both the sex worker is offering and the client is seeking while keeping the rest private.

MPCs can also be used as another approach to privacy-preserving voting and may facilitate new mechanisms for privacy-preserving community consensus. For example, let's say we are a data collective, deciding on permissible use cases for our collective data and what organizations may be granted access to our data. We could each grant consent to differing use cases, ideally through some form of accessible interface. Access is only granted when some agreed-upon consensus is reached, and otherwise revoked. The identities of who voted, and what their vote was, are kept private.

I am really excited about the possibilities that secure multi-party computation can enable. I think that it can be a foundational building block for creating infrastructure that can be *both transparent and privacy-preserving*—two properties that are typically seen as mutually exclusive trade-offs.

I also see MPCs as a potential solution to the prisoner's dilemma, a way to say, "I'm willing if you're willing" in situations where *the act of revealing information can change the circumstances.* This is especially helpful in cases where revealing information first can result in a penalty or disadvantage, making MPC a candidate for facilitating new mechanisms for collective trust.

Fully Homomorphic Encryption

Fully homomorphic encryption (FHE) is an advanced encryption scheme that allows computations to be performed on encrypted data without decrypting it first. In traditional encryption schemes, you would need to decrypt data to perform computations, creating potential vulnerabilities in data security. FHE allows data to remain private, secure, and confidential, even throughout computations run by third-party services.

FHE has been lauded by some as the holy grail of cryptography. It allows *use and access* of services that may need personal data to provide the service, and since the server is running on encrypted data, that personal data is kept private throughout the entire process.

I'm fairly excited for FHE as a possible tool for enabling free use and access in a digital ecosystem while still maintaining privacy. FHE is a very powerful privacy tool. It's still quite new and pretty computationally expensive. While in theory any service can be transformed into one that is privacy-preserving using FHE, it may not *yet* be practical given the computational resources it uses.

Let's say we are a data collective with vulnerable members, such as sex workers, community organizers, or journalists, and we store all of our data in an encrypted database. We may allow other entities, such as an advocacy group or academic research organization, to run analysis on our data. With FHE, they can access and use our data, ask questions, run computations, and receive answers, without owning or even viewing anything that may reveal personally identifiable information about any of our members. Since both the data and the computations are within an encrypted ecosystem, no individual details are exposed or compromised throughout the process.

To summarize, zero-knowledge proofs allow cryptographic proof of knowledge without revealing the information itself; multi-party computation enables joint computation without revealing private inputs; and fully homomorphic encryption enables secure computations on encrypted data without the need for decryption during the computation process.

These privacy and security strategies collectively contribute to a greater holistic ecosystem. ZK, MPC, and FHE may be combined with other privacy, security, and cryptography strategies, such as verified credentials, decentralized identity, differential privacy, federated learning, other forms of encryption, and many more **privacy-enhancing technologies (PETs)**. Furthermore, even with these emerging privacy and cryptographic frameworks, technical approaches still exist within greater social, legal, and cultural contexts that require advocacy and consideration just as well.

Data Collectives

In addition to the privacy focus is an emphasis on the structure of *data collectives*: data co-ops, data trusts, data coalitions, data unions, data commons, etc.

How do these different infrastructures inform practices around governance, membership, access, and usage, and vice versa?

For example, a data trust may imply the existence of trustees and beneficiaries. A data co-op may imply a governance structure of one member, one vote. A data coalition may imply coalition building and possible structures around membership, such as whether membership can overlap and, if so, how. A data commons may imply open access and collective stewardship.

Enabled by cryptographic tools, how can data collectives promote collective consent mechanisms and community power?

Interfaces for Data Consent

I've also been thinking a lot about **interfaces for data consent** and possible infrastructures—social, legal, technical, or otherwise—for facilitation and stewardship of these collective data frameworks. I'm inspired by Creative Commons licensing; copyleft and open source licensing, such as GPL3; GDPR cookie consent interfaces; "consent profiles;" robots.txt protocols; and co-design approaches in community research.

Creative Commons Licensing

I feel particularly inspired by Creative Commons licenses and the way they provide legal, social, and cultural infrastructure around access, distribution, and use. While few creatives have the resources to fully enforce the terms of the licenses, the existence of this framework makes it possible for creators to both protect as well as freely share creative work, media, and materials. I'm also supportive of the framework of the **commons** and the implication that these works are community resources, meant to be accessible to all.

It's an interesting comparison and contrast to the histories of copyright as well as patenting. While initially developed to encourage the sharing of new work and inventions, copyright and patent law increasingly became influenced by lobbying efforts to protect and privatize intellectual property on behalf of corporations (see: Disney's "Mickey Mouse Protection Act").

Revoking Data Consent

The case of data deletion and consent revocation is interesting and potentially hairy. How can we delete data that has already been used to train a model? Can we prove or provide guarantees that the model has been retrained without using one's data? What other ways exist to revoke access to data that has already been used?

Outside of cryptographic tools, there exists a handful of privacy services that claim to delete your personal data from data brokers: Incogni, a "personal information removal service;" Onerep, a "fully

automated privacy service;" and Consumer Reports' Permission Slip, an "app to take back control of your data."

It would be interesting to be able to visualize the flow of data collected, accessed, and used. When I accept a cookie, what am I really agreeing to? Which data brokers now have my data, and who are they selling and trading it to? Who now uses the information I've just granted permissions to, knowingly or unknowingly?

With the framework of data consent, there exists a distinction between consent and **informed consent**. I'm curious how we could build better tools, interfaces, and infrastructure for granting and revoking informed consent in digital culture.

Models Beyond Ownership

What's **beyond ownership and control**? I dream of seeing frameworks of ***access, usability, and responsibility*** *over ownership* as well as ***stewardship, consent, and agency*** *over control.*

While I appreciate models around "collective ownership" or "letting users control their data," I don't see them as the ultimate end goal. For example, the concept of land ownership did not exist in Indigenous culture; the framework of private property is a colonial import.

Community infrastructures that inform these frameworks may include public transit, public parks, the library, and public archives—infrastructures where *ownership* of land, vehicles, books, and other objects *is not necessary to enable use and access* of common resources.

How might these models extend to other goods and services, such as housing, food, or education?

Community Research and Co-Design

For a more robust implementation of these applications, I'd encourage designing these technologies by and with communities. To actually serve vulnerable populations, rather than only offer another instance of tech solutionism, awareness of the limitations of pure technology is essential. At what point can technical tools for privacy and cryptography help vulnerable communities, and at what point are the remaining challenges outside the scope of technology? What cultural, social, legal, and other infrastructural barriers exist, and how can they be addressed in tandem?

My hope is that technologists can do our due diligence in community research, identifying the specific needs, wants, challenges, and priorities of vulnerable populations in an intersectional, holistic, and interdependent way. Ideally, co-designing these processes and implementing these technologies from the ground up is a collaborative approach with the communities most directly impacted and most in need of privacy protections. In many ways, the process of building and implementing technology is inherently political. The way we build these tools and frameworks, and who we build them for and with, will have lasting cultural and social impacts downstream.

About the Author

Riley Wong is the Principal of Emergent Research, where they investigate community governance and cooperative infrastructures, with emphasis on decolonial and international perspectives; and privacy-preserving data-governance, i.e., cryptography-enabled data collectives for sex workers and other vulnerable populations.

They have backgrounds in machine learning at Google, investigative journalism at ProPublica, healthcare research at Penn Medicine, and QTBIPOC community organizing with the Audre Lorde Project.

Their research is rooted in community practice and guided by collaborative work in solidarity economies, public goods and the commons, peer-to-peer systems, interdependence, and emergence, and they hope to find new possibilities for collective power, agency, and consent when privacy tools are integrated with a community-first approach.

They can be found at rileynwong.com and emergentresearch.net.

About the Ash Center

The Mission of the Roy and Lila Ash Center for Democratic Governance and Innovation at is to develop ideas and foster practices for equal and inclusive, multi-racial and multi-ethnic democracy and self-government.

About the Second Interdisciplinary Workshop on Reimagining Democracy

This essay was adopted from a presentation given at the Second Interdisciplinary Workshop on Reimagining Democracy held on the campus of Harvard Kennedy School in December 2023. Convened with support from the Ash Center for Democratic Governance and Innovation and the Belfer Center for Science and International Affairs, the conference was intended to bring together a diverse set of thinkers and practitioners to talk about how democracy might be reimagined for the twenty-first century.

This essay is one in a series published by the Ash Center for Democratic Governance and Innovation at Harvard University's John F. Kennedy School of Government. The views expressed in this essay are those of the author and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. The papers in this series are intended to elicit feedback and to encourage debate on important public policy challenges.

This paper is copyrighted by the author(s). It cannot be reproduced or reused without permission. Pursuant to the Ash Center's Open Access Policy, this paper is available to the public at ash.harvard.edu free of charge.

A PUBLICATION OF THE

Ash Center for Democratic Governance and Innovation
Harvard Kennedy School
79 John F. Kennedy Street
Cambridge, MA 02138
617-495-0557
ash.harvard.edu

A PUBLICATION OF THE

Ash Center for Democratic Governance and Innovation
Harvard Kennedy School
79 John F. Kennedy Street
Cambridge, MA 02138

617-495-0557
ash.harvard.edu

HARVARD Kennedy School
ASH CENTER
for Democratic Governance
and Innovation