Federal Register Notice 87 FR 35250, https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies, June 9, 2022

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

OpenMined

RFI Response: Privacy-Enhancing Technologies

Office of Science and Technology Policy
White House
1650 Pennsylvania Avenue NW
Washington, DC 20502

> Re: Request for Information on Advancing Privacy-Enhancing Technologies (87 FR 35250)

We appreciate the opportunity to respond to this Request for Information on Advancing Privacy-Enhancing Technologies.

OpenMined is a nonprofit organization that develops leading free and open-source software allowing data scientists and A.I. or machine learning practitioners to use advanced privacy enhancing technologies (PETs). We educate thousands of professionals about PETs; organize a technical and professional community; and deploy boundary-pushing implementations of PETs with institutions including Twitter and the UN PET Lab. The best place to learn more about our work is our blog.

Our mission is to unlock the stunning potential of non-public data to advance social good, while simultaneously helping engineers, institutions, and society navigate the complex hazards presented by these technologies. Neither of these aims, we believe, can be effectively pursued without the other.

1. **PETs are infrastructure. Network effects and technical lock-in make it critical for the government to use and support free and open-source PETs standards. (Questions 1, 2, 5, 7, 8)**

Infrastructure and the public interest

PETs are more than useful tools. They are vital infrastructure for tomorrow's digital networks. Therefore we believe the most important way the government can ensure that PETs work in the public's interest is to **favor free and open-source PET standards through policy, resources, and official action**.

When the modern internet coalesced in the early 1990s, the work of farsighted policymakers helped ensure that it was built on a foundation of free and open source software, such as the

TCP protocol from the earlier ARPANET. This stroke of leadership was critical to bringing about a true, open “world wide web”. Otherwise, the internet would have emerged in the form of walled gardens dominated by rent-seeking private actors, like AOL and the major telecommunications companies. The dynamic startup culture of the 1990s and 2000s might not have emerged. Or, the revolution of an open internet might have started elsewhere, with the United States failing to lead it.

Network effects, technical lock-in, and competition

In the coming years, PETs will increasingly supplement legacy data transmission protocols with more precise and targeted methods of information sharing. This will represent a tectonic shift in network architecture, comparable to the emergence of the early internet.

As in the 1990s, network effects will create technical and economic pressure toward the consolidation of a *single* interconnected network (or at most a few), through which information may be shared using common software protocols. This means that whatever PET-enabled networks seize the first-mover advantages could quickly achieve global scale and lasting strategic dominance.

Let us walk through a hypothetical example of how this could happen. Suppose numerous important holders of sensitive medical data, like hospitals, decide to make their combined data available to researchers through advanced privacy preserving methods. To this end, they set up access and permissions to use a very large combined dataset. Their initiative is successful: it dramatically accelerates researchers’ progress, without disturbing upstream privacy, confidentiality, or intellectual property interests.

This could result in that network growing exponentially. Why? Since data is more valuable when combined with complementary data, a leading PET-enhanced data sharing network would be able to pay the market’s best prices for additional contributions of information, compounding its advantage. The network’s utility would likely expand beyond its initial ambit into other fields with no obvious limit. Access to it would become indispensable not only to medical researchers, but to practitioners of a wide variety of A.I. and machine learning disciplines, and the businesses that rely on their work (i.e., most large businesses).

In this way, the maturation of PETs could have extremely negative consequences and lasting consequences regarding competition. A dominant network dependent on proprietary PET standards would give private actors irresistible opportunities to extract rents and impose switching costs. Exclusive reliance on free and open source PET protocols is therefore a necessary (albeit not sufficient) requirement to protect the public and the economy from foreseeable harm.

The government should use its substantial leverage to send the industry down the path of free and open-source PETs standards, as described below.

Supporting open-source PETs

The government should support free and open source PETs through (a) policy, (b) resources, and (c) official action to speed adoption. By throwing its considerable weight behind these efforts, the government can help ensure that the dominant PET-enabled networks that emerge in coming years will not rely on proprietary standards.

Regarding *policy* and *resourcing*, we broadly support the **Promoting Digital Privacy Technologies Act** (H.R. 847 and S. 224) and the related **DATA Privacy Act** (H.R. 5807 and S.3065), both of which would accelerate the development (and therefore adoption) of PETs. H.R. 847 contains language at Section 3(7) directing the National Science Foundation (NSF) to support "development of freely available privacy enhancing technology software libraries, platforms, and applications". We applaud this clear emphasis. **This and similar legislation should be complemented by a firm operational understanding of PETs' competition implications, by decision-makers across all Federal agencies.** To this end we believe that efficient and focused education initiatives directed toward key Federal officials is warranted. OpenMined would be happy to comment further on this.

The **Secure Research Data Network Act** (S. 4200) is also an encouraging initiative. This bill would create responsible Federal data stewards helping researchers benefit from Federal data. Such stewards could help shape best practices, increasing the chances that important PETs networks are operated in the public interest. At Section 3(D)(2), the proposal provides that such networks must be built with free and open source software. This provision is of vital importance for reasons we have articulated.

Regarding *official actions to speed adoption*, much can be undertaken in advance of or in addition to legislation. For example it is already possible for any and all agencies to host dedicated federated data networks providing safe access to the information they steward. Early examples include OpenMined's present work on pilot initiatives with the U.S. Census Bureau. In addition to replicating and expanding projects of this kind, the government should support academic work in any discipline that demonstrates the utility of the newly-accessible data. For example, NSF grants could favor applications that express an intention to explore or leverage such data.

2. **PETs serve democracy by enabling transparency into important networks. (Questions 1, 2, 3, 4, 9)**

PETs can unlock maximum transparency into technical systems that have strategic or public importance – such as those run by governments or large technology companies – while respecting privacy, security, or intellectual property concerns.

This is one of PETs' most important implications for justice, economic fairness, and the future of democratic institutions.

Examples of systems that PETs can make more transparent include:

- social media companies' algorithms, which may fuel radicalism, polarization, or foreign political influence
- new digital networks using decentralized technologies like blockchains, which may generate complex, subtle, and hard-to-foresee distortions of social, political, and economic life
- enforcement patterns at regulatory or law enforcement agencies, which may exhibit complex unfairness along racial, gender, or other protected qualities

Today, only the parties that own and operate these systems have the information necessary to understand and identify their complex public implications. Yet, analyzing and understanding these implications is at best outside the expertise, and at worst against the incentives of such parties. It is therefore critical to develop techniques allowing third parties like researchers, civil society organizations, and government regulators, to study and understand the patterns of activity in these networks.

PETs are our sharpest tools for advancing the frontiers of transparency while respecting the privacy, security, and intellectual property concerns that weigh in favor of opacity. Thus, investing in both fundamental PETs techniques and ambitious applied projects is necessary if our technical systems are to operate in accordance with open and democratic values.

Supporting ambitious demonstrations of techniques that can address algorithmic bias and related problems

OpenMined is conducting a pilot with Twitter that advances and demonstrates PETs' transparency-unlocking possibilities, and could serve as a precedent for other social media companies.

The project uses a large, 1-billion-tweet private dataset from Twitter, which researchers could not access through the ordinary Twitter API. It creates a synthetic version of that data, and makes it available to researchers to perform computation on it. Researchers can see the results of their computations using an adversarial differential privacy system. This system tracks privacy budgets on the level of individual Twitter users, not as an aggregate dataset. This enables researchers to obtain much more detailed information than conventional differential privacy deployments, while still ensuring that they cannot reconstruct individual information that would have privacy implications.

The project is experimental. It remains to be seen, for example, how useful researchers' findings will be concerning Twitter's algorithms' influence on political discourse. However, if successful, it will shift industry norms in terms of how much transparency platforms should provide into their algorithms. **Platforms offering researchers less access than the standards set by this and future programs will not be able to claim the fig leaves of privacy, security, or trade secrets.**

Applied projects like this are perhaps the most important activity that the government can support to nudge the technology industry towards better transparency practices.

Regulating decentralized networks with PET-enabled safe havens

In the short term, most of the transparency-related opportunities for PETs concern building windows into centralized networks. However, the government should also be aware of the medium-term relevance of PETs to the regulation of decentralized networks.

Today, the largest blockchain networks are anonymous, but public. This permits a surprisingly high degree of traceability for regulators. However, some less-popular decentralized networks are truly resistant to tracing, and therefore potentially problematic for law enforcement. Such hard-to-trace networks can be expected to proliferate and/or grow in sophistication. It is not yet clear what strategy the government can or should take in this area.

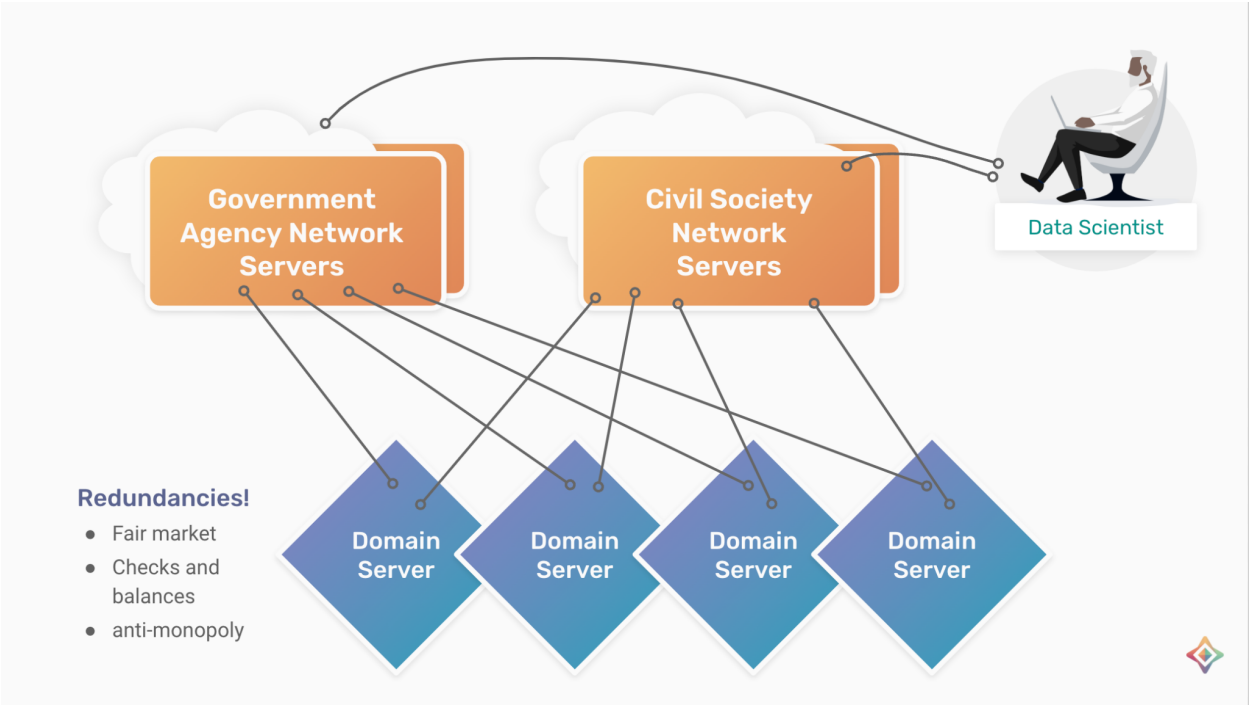
PETs create interesting, although still speculative possibilities. First, it is important to acknowledge there are legitimate and potentially beneficial applications for hard-to-trace decentralized networks, which should be weighed against their propensity for misuse. For example, consider communities wishing to share resources in oppressive regimes; or legitimate activists seeking to coordinate mutual aid networks without exposing themselves to social or professional retribution. The government might permit hard-to-trace networks to operate for such legitimate purposes by crafting PET-enabled safe-harbor rules. For example, otherwise-untraceable networks could be licensed if their architecture enabled transparency sufficient to rule out illicit uses while preserving essential privacy features.

Helping public institutions to develop best practices for using PETs

Government agencies and regulators themselves often harm the public through structural bias. Letting researchers study internal data from agencies such as the IRS, Census, or DOJ could uncover such bias and improve the government's ability to serve the public. The **Secure Research Data Network Act** (S. 4200) would be a welcome step forward in this regard.

OpenMined has partnered with Census, IRS, and the UN PET Lab to explore the utility of PETs for agency use. Census and IRS manage sensitive data of huge public value, which our systems help them open to new uses. Similarly, our work with the UN PET Lab lets trade authorities of competing nations reconcile complex data on imports and exports, helping them identify accounting discrepancies (and accelerate trade), without needing to share sensitive data with non-trusted counterparts.

Going a step further, our systems enable unique forms of collaboration between government and civil society. Responsible civil society organizations can be directly incorporated into the networks through which researchers access sensitive data; this serves as a check and balance on the decisions of agency network administrators.



Analogously, OpenMined's AdAstra project has established a federated learning network across private data servers located at over 100 universities and research institutions. This is an unprecedented effort to unify disparate internal review processes for accessing research; while also building a system of unhindered privacy-preserving access so that researchers can work from exponentially larger datasets.

Establishing PETs in the public and nonprofit sector has long-term strategic and economic implications. Agencies, more than private companies, are genuinely accountable to the public interest. This means that, with good leadership, their incentives are better aligned toward developing responsible uses of PETs. Moreover, dovetailing with our argument in Section 1, PET networks operated by government agencies could grow exponentially and become strategically important.

It is much better for the stewards of strategic networks to be agencies, which are accountable to the U.S. public, than private companies accountable to shareholders and potentially foreign interests.

3. Encouraging bottom-up and shared data governance. (Questions 5, 10)

The architecture of data sharing will change fundamentally in the coming years. Optimistically, this shift will involve the establishment of a new class of innovative, fiduciary-style institutions that manage data on behalf of interest-aligned consumers, businesses, and others. Broadly similar versions of this idea give different names to such hypothetical institutions, including data coalitions, data trusts, data cooperatives, mediators of individual data, middleware, and more.

These evolving ideas hold medium-term promise. Trustworthy institutions could, for example, help individual consumers aggregate their interests and mitigate the take-it-or-leave-it dynamics that characterize their present interactions with dominant platforms. PETs are an important part of this vision: they will enable such institutions to exert much more meaningful control over data.

Requirements that digital service providers not monopolize their customers' data – similar to those imposed on IoT providers by Chapters II-III of the EU Data Act – would help support the emergence of such PET-enabled data governance, by preventing data from being locked in to dominant platforms.

Conclusion

The government should view PETs as basic infrastructure. Supporting free and open-source standards reduces the danger of harmful monopolies or rent-extraction that could harm the public and hinder U.S. competitiveness.

PETs have an important role to play in restoring and maintaining the health of democracy by enabling secure transparency into important institutions, from government agencies and technology companies to decentralized networks.

PETs may also unlock possibilities for promising forms of shared data governance that could empower consumers and others.

The government can and should support positive progress by supporting legislation, encouraging research, and actively incubating more and better-funded experiments. Decision-makers across federal agencies should be educated about PETs' strategic considerations. To accelerate adoption, federal funds should focus on the demand side in addition to the supply side, e.g., training and supporting researchers whose work makes use of public data made available through PETs.

OpenMined looks forward to actively assisting this work. Please do not hesitate to contact us – we would be delighted to continue the conversation.

Respectfully Submitted,

Andrew Trask
Leader, OpenMined

Matt Prewitt
Advisor, OpenMined
President, RadicalxChange Foundation