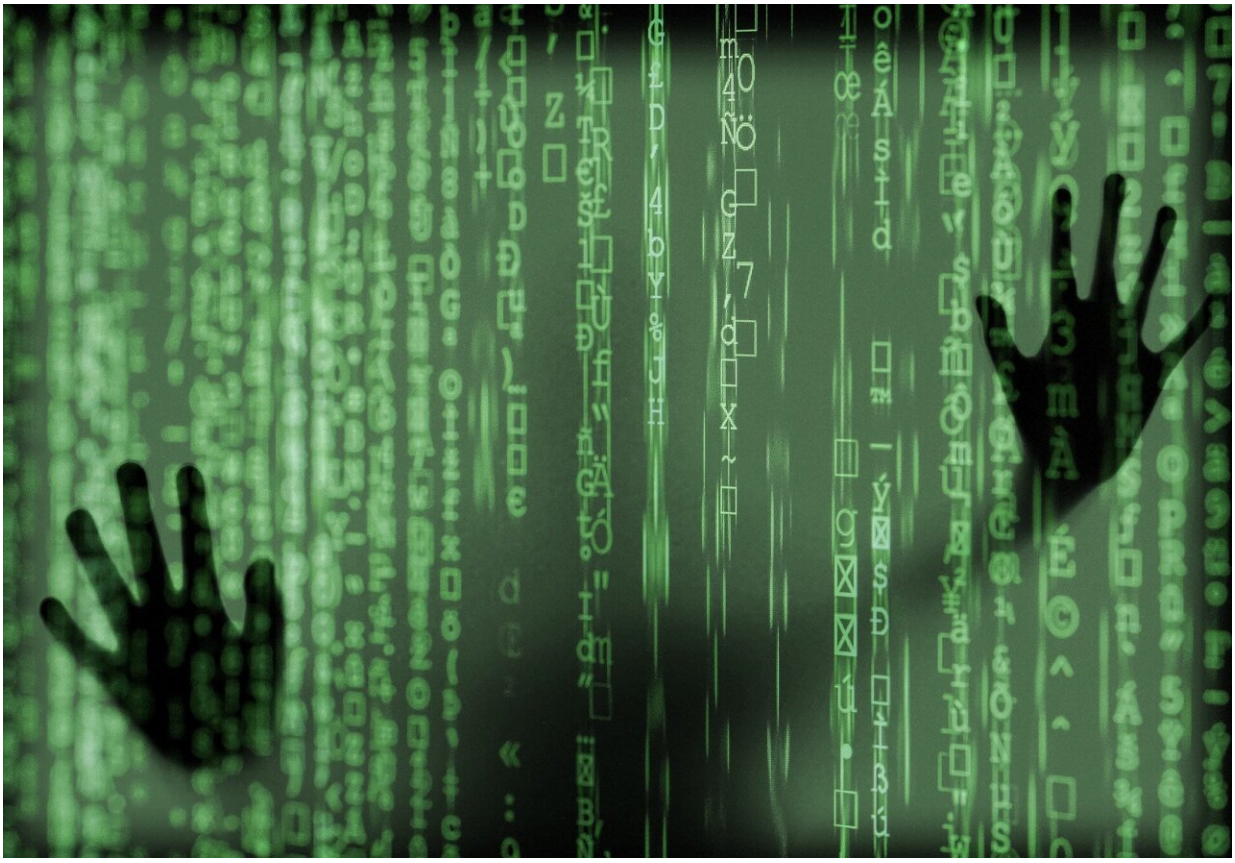
# Why the search for a privacy-preserving data sharing mechanism is failing

June 2 2022, by Tanya Petersen



Credit: Pixabay/CC0 Public Domain

From banking to communication our modern, daily lives are driven by data with ongoing concerns over privacy. Now, a new EPFL paper published in *Nature Computational Science* argues that many promises

made around privacy-preserving mechanisms will never be fulfilled and that we need to accept these inherent limits and not chase the impossible.

Data-driven innovation in the form of personalized medicine, better public services or, for example, greener and more efficient industrial production promises to bring enormous benefits for people and our planet and widespread access to data is considered essential to drive this future. Yet, aggressive data collection and analysis practices raise the alarm over societal values and fundamental rights.

As a result, how to widen access to data while safeguarding the confidentiality of sensitive, personal information has become one of the most prevalent challenges in unleashing the potential of data-driven technologies and a new paper from EPFL's Security and Privacy Engineering Lab (SPRING) in the School of Comupter and Communication Sciences argues that the promise that any data use is solvable under both good utility and privacy is akin to chasing rainbows.

Head of the SPRING Lab and co-author of the paper, Assistant Professor Carmela Troncoso, says that there are two traditional approaches to preserving privacy, "There is the path of using privacy preserving cryptography, processing the data in a decrypted domain and getting a result. But the limitation is the need to design very targeted algorithms and not just undertake generic computations."

The problem with this type of privacy-preserving technology, the paper argues, is that they don't solve one of the key problems most relevant to practitioners: how to share high-quality individual-level data in a manner that preserves privacy but allows analysts to extract a dataset's full value in a highly flexible manner.

The second avenue that attempts to solve this challenge is the anonymization of data—that is, the removal of names, locations and

postcodes but, Troncoso argues, often the problem is the data itself. "There is a famous Netflix example where the company decided to release datasets and run a public competition to produce better 'recommendation' algorithms. It removed the names of clients but when researchers compared movie ratings to other platforms where people rate movies, they were able to de-anonymize people."

More recently, synthetic data has emerged as a new anonymization technique however the paper suggests that, in contrast to the promises made by its proponents, it is subject to the same privacy/utility trade-offs as the traditional anonymization of data. "As we say in our paper researchers and practitioners should accept the inherent trade-off between high flexibility in data utility and strong guarantees around privacy," said Theresa Stadler, Doctoral Assistant in the SPRING Lab and the paper's co-author.

"This may well mean that the scope of data-driven applications needs to be reduced and data holders will need to make explicit choices about the data sharing approach most suitable to their use case," Stadler continued.

Another key message of the paper is the idea of a slower, more controlled release of technology. Today, ultra-fast deployment is the norm with a "we'll fix it later" mentality if things go wrong, an approach that Troncoso believes is very dangerous, "We need to start accepting that there are limits. Do we really want to continue this data driven free for all where there is no privacy and with big impacts on democracy? It's like Groundhog Day, we've been talking about this for 20 years and the same thing is now happening with machine learning. We put algorithms out there, they are biased and the hope is that later they will be fixed. But what if they can't be fixed?"

Yet narrow functionality and high privacy is not the business model of the tech giants and Troncoso urges that all of us think more carefully

about how they address this critical issue.

"A lot of the things that Google and Apple do is essentially whitewash their harmful practices and close the market. For example, Apple doesn't let apps collect information but collects the data itself in a so called 'privacy preserving' way, then sells it on. What we are saying is that there is no privacy preserving way. The question is 'did the technology prevent harm from the system or did it just make the system equally harmful'? Privacy in itself is not a goal, privacy is a means with which to protect ourselves," Troncoso concludes.

**More information:** Theresa Stadler et al, Why the search for a privacy-preserving data sharing mechanism is failing, *Nature Computational Science* (2022). DOI: 10.1038/s43588-022-00236-x

Provided by Ecole Polytechnique Federale de Lausanne