Owning the Wind

Henry Lieberman
Computer Science and
Artificial Intelligence Lab (CSAIL)
Massachusetts Institute of Technology (MIT)
Cambridge, MA 02139 USA
lieber@media.mit.edu

Each submission will be assigned a unique DOI string to be included here.

Abstract

Consumer and civil rights advocates are now debating how to assure protection for individuals from abuse of data by corporations and governments. The urgency of such protections has drastically increased in the last few years due to "Big Data" and Artificial Intelligence programs which make increasing use of personal data. The debate is often framed as being about "ownership of data", with the idea that "the user owns their own data" as being the foundation for protection of individuals.

This isn't wrong, but ownership might not be the best framework for thinking about the problem. Instead, an alternative view is that data is *entrusted* to organizations that use it, and the obligation to make fair and ethical use of the data continues permanently. The best protection for individuals is to ensure that organizations *cooperate* rather than *compete*, with their constituents.

Author Keywords

Intellectual Property. European General Data Protection Regulations.

ACM Classification Keywords
K.4 COMPUTERS AND SOCIETY
K.5 LEGAL ASPECTS OF COMPUTING

Introduction

Once, during the 19th century genocide of Native Americans, there was a story about a US general who was trying to negotiate a "treaty" to buy land from a tribe of Native Americans. The chief was able to understand some English. The general tried repeatedly to ask the chief how much he wanted in order to purchase the land, but it was clear that the chief wasn't understanding the question. The chief turned to one of his colleagues who had better English, and asked, in his native language, if he could translate what the general was asking. The interpreter launched into a long-winded explanation in the indigenous language. When he reached the end, all the native Americans burst out in raucous laughter. It took a few minutes for it all to die down.

Then the chief spoke one sentence, still barely able to control his laughter. That sentence sent them all back into paroxysms of laughter all over again. After it died down again, the general turned to the interpreter.

"What did he say?"

"He said, 'Next, you're going to tell us you want to buy the wind.'"

Among advocates for civil liberties and consumer protection, there's a lot of discussion these days about ownership of data. Personal assistant software and "big data" analytics hold a lot of promise for providing services that improve people's lives, but they inherently require users to share personal and potentially sensitive data. If you want the doctor to help you, you have to take off your clothes.

Rightly, there's concern that personal data might be used to the user's detriment by malicious companies and governments. People have been denied health insurance or home loans because the provider was acting on personal information that they shouldn't have had. There's been automated discrimination against racial and ethnic groups. Undoubtedly, people have lost money, jobs, and probably lives.

It's worth trying to figure out how society can protect individuals against such abuses. These are the kind of concerns that motivate the European General Data Protection Regulations [2], and the recent calls for a Right to Explanation [5] of decisions made by Artificial Intelligence programs. These efforts are laudatory.

Such regulations may be the best we can do under the current circumstances. But they have a number of problems that are currently being debated. How do we tell when data is really "personal"? How do we prevent "triangulation" of personal data that can be inferred from public data and/or other personal data? How can these rules be enforced? Will they really protect users?

This debate is often framed as, "Who owns the data?". The answer is, and must be, "The user owns their own data" [4]. The idea is, that in order to protect the user from abuses, the user must be in ultimate control of the data. That's right.

But is ownership the right way to frame the issue? Our society seems to be able to deal reasonably well with ownership of physical objects, but extending this metaphor to intangible concepts like data is proving problematic. The endless contention around so-called "intellectual property" shows that there is no real consensus about the notion of "owning" well, what, exactly? Ideas? "Embodiments"? Designs? Processes?

Ask the intellectuals like me, what they think about intellectual property. You'll find that few really mount a strong defense of the idea. Few really think that the current systems of patents and copyrights really work in their interests. Licensing and sale of intellectual property has never been a steady, reliable, and profitable business model for creative people, or for universities, except for a very few in very unusual circumstances [1] [7]. However, it is indeed a steady, reliable and profitable business model for armies of lawyers, corporations, and players like patent trolls.

I have a few patents myself. They've been of worthless, or even negative value. I do, however, have a lucrative sideline in being an expert witness in IP cases. I feel somewhat guilty about it. My only defense of myself here is that I almost always work for the defense attorneys, that is, to reduce the amount of (enforceable) intellectual property in the world.

Is society really served by extending the dysfunctional concept of IP to data? Does anyone really think that "selling your data" will bring any kind of substantial benefit to the end users? Or will it, like IP today, be more welfare for lawyers?

Love [6], in a study of the profitability of university patents (which probably are more credible than lone-inventor and legally strategic corporate patents), calculated that the Return on Investment in patent activities was a *negative* 3%. He did not find any evidence that the prospect of patents provided any tangible incentive for researchers to conduct more or better research. 57% did not even know whether their university policy entitled them to share patent revenues.

If you own something, that means someone else cannot use it without your permission. It is not fair if your data is used without your knowledge and permission, or for purposes that you do not approve of. The dilemmas come when we consider questions of under what circumstances that knowledge and permission are solicited and granted.

Companies and governments would like to perform these transactions in a "clean" manner. That is, they ask you for your permission to use your data, through asking you to sign a contract. They specify what they can and cannot do with it. Maybe you get paid for it. If so, it's probably not much. More likely, you don't. They make you agree by standing in the way of something else you want to do. Sometimes they do so by obfuscatory legal language and vague, unverifiable promises that they will use it "to improve service".

Then the data is "theirs", not yours. They are not responsible for any consequences after that. That is, permission can be transferred, possibly with an exchange of money. The problem is that knowledge and permission may have consequences. You can't specify the all the consequences in advance. Often, the

"buyer" will know full well that there will be negative consequences for you that you might not realize, but they won't tell you, and nobody can make them tell you. If you sign, it's your fault. Ask the Native Americans how well that worked out for them.

In the end, ownership of data might prove as oxymoronic as ownership of the wind.

I submit that you can't buy or sell personal data. Any more than you can buy or sell people. What you can do is *share* data with someone else. Whenever you share data, you *entrust* your data to whomever you share it with. And that trust is sacred. The responsibility to use someone else's data in ethical and mutually beneficial ways cannot be discharged. Not by money, not by contracts, not by promises. It remains throughout any subsequent use of the data.

It's like if you tell a friend a personal secret. I might give you specific instructions about whom to tell or not, or what use you can make of knowing my secret. But I can't specify every possible situation. It's your responsibility to treat that information, to the best of your ability, in the way you think I would want you to, or the way you would treat it if it was your secret rather than mine. That responsibility lasts forever. That's what a good friend does. Otherwise you're simply not worthy of my trust.

Companies and governments won't like this answer. It precludes the possibility of a "clean" transaction. Regulators, privacy advocates, and consumer advocates aren't going to like this idea either. It also precludes the possibility of "clean" rules and regulations governing what companies and government agencies can do. But I think that's the only real answer.

It comes down to value alignment between the organization and the individuals. Are the companies trying to work with their customers, or against them? Are they trying to cooperate with the customers and with their employees, or compete against them? Are governments working for their citizens, or against them? Companies say, "The customer comes first." If so, now is the time to prove it. Governments say, "We're working on behalf of the citizens". If so, now is the time to prove it.

As we all know, in today's real world, the interests of organizations and individuals aren't always aligned. In the case of companies, the (American) idea of "fiduciary responsibility", says that the values of the company are *not* the interests of the customers and the employees, but of the stockholders. It's really not, "The customer comes first", but "The customer's money comes first". Despotic governments are upfront about the fact that their power "trumps" the interests of the citizens. Democratic governments are better, but here too, the government asserts power over individuals in ways that preserve and extend the government's control, even if that is not in the interest of the citizens. Democratic governments do have mechanisms like elections that can curb the worst abuses, and bring values closer into alignment, though those mechanisms are slow and unreliable as protections. So these organizations are literally and inherently, untrustworthy. Their values do not align, as the values of two friends do.

That being the case, many people think we just have to accept the misalignment of values between organizations and individuals. Organizations and individuals will always "compete" with each other. The best we can then do, is to establish rules for a "fair fight" and hope for the best. That's where regulations like the European data protection regulations come in. That's where companies that try to do their best to establish user-friendly privacy policies come in.

In the short term and in the medium term, that is, indeed, probably the best we can do. So I do want to thank those who do try to improve end-user data policies, both in companies and in governments. But it's like the Marquis of Queensberry rules for boxing. They make the fight fairer, but it's inevitable that somebody's still going to get hurt. And in a contest between organizations and individuals, who do you think is going to end up face down on the mat? In a contest between machines and people, who do you think is going to end up face down on the mat?

So I think the real long-term solution is that we have to begin the process of trying to transform our economic and political organizations so that they operate in the interests of individuals they are intended to serve. Then, and only then, will they be trustworthy enough to entrust our personal data to. Organizations will have to learn to cooperate with their constituents, not compete with them.

The good news is that I think it will be very possible to achieve it. Though maybe not quickly. And, perhaps surprisingly, it'll be the very technologies that we're worried might threaten people, like Artificial Intelligence, that'll enable us to get there. So, while we do have to keep the today's abuses of data from getting out of hand, we also need to be careful to make sure that regulations and fear won't kill the golden goose that will solve the fundamental problem.

That fundamental problem is *scarcity* [7]. Material scarcity is the root cause of competition between people. All the systems we have today, economic and political, are based on competition, and it's, literally, killing us. Scarcity is the reason why organizations compete with their constituents. Scarcity is the reason why today's governments are based on power relationships. If you think I'm fundamentally arguing against the very idea of Capitalism, you'd be right. If you think I'm fundamentally arguing against the way all governments are set up today, you'd also be right. These are radical ideas.

But the transformation from an industrial society to a data-driven society will be just as great as the societal transformations that brought us Capitalism and western Democracy in the first place, starting in the 18th century. So we should not be afraid to consider, for the long term, solutions that are just as radical as those were in their day. We shouldn't dismiss any talk of radical solutions with, "That's so far in the future we won't talk about it now". The urgency for today is to understand the directions the solutions have to take, and get started. Even if it's 100 years off, let's do 1% of it next year.

I don't expect to convince you immediately. Christopher Fry and I have written a book, entitled *"Why Can't We All Just Get Along?"* [3], that presents a vision of a post-scarcity future, and concrete proposals for new cooperative economic and political structures. It doesn't

talk very much specifically about data sharing. But we believe it takes care of the fundamental problem. Read it, and then tell us if you think data sharing would still be such an enormous problem in the world we describe. We think it won't be.

So it would be insane to turn data into a commodity, to be bought and sold. Instead, the solution is to turn commodities into data. Personal manufacturing (3D printing) turns hardware into software. Software is replicable, so physical objects can be replicable. Artificial Intelligence ends scarcity of labor. We can then figure out how to transform organizations so they cooperate with citizens, not compete with them. When we do, bickering about "who owns the data", will go away.

The chief was right. You can't own the wind. Nobody can. But if you work with the wind instead of trying to constrain it, it can fill your sails and move you forward.

References

1. Michele Boldrin and David Levine, The Case Against Intellectual Property, American Economic Review. Apr 2002, Vol. 92, No. 2: Pages 209-212.
2. European Union, European Union General Data Protection Regulation, http://www.eugdpr.org, 2018.
3. Christopher Fry and Henry Lieberman, Why Can't We All Just Get Along?, http://www.whycantwe.org/, 2018.
4. Daniel Greenwood, et. al. The New Deal on Data: A Framework for Institutional Controls, in *Privacy, Big Data, and the Public Good*, J. Lane et. al., eds, Cambridge University Press, 2014.
5. Bryce Goodman and Seth Flaxman, European Union regulations on algorithmic decision-making and a "right to explanation", International Conference on Machine Learning, Workshop on Human Interpretability, New York, 2016. https://arxiv.org/abs/1606.08813
6. Brian J. Love, Do University Patents Pay Off?, Yale Journal of Law and Technology, 16(2), 2014.

   http://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=1105&context=yjolt
7. Eldar Shafir and Sendhil Mullainathan, *Scarcity*, Macmillan, 2013.
8. Siva Vaidhyanathan, Copyrights and Copywrongs: The Rise of Intellectual Property and how it threatens creativity, NYU Press, 2001.