Beyond Data Ownership

Beyond Data Ownership to Information Sharing

The question of who owns our data on the Internet is a challenging problem. It can also be a red herring, distracting us from building the next generation of online services.

The term “ownership” simply brings too much baggage from the physical world, suggesting a win-lose, us-verses-them mentality that retards the development of rich, powerful services based on shared information.

I’m a member – and a big fan – of Steve Holcombe’s “Data Ownership in the Cloud” LinkedIn group and I love the efforts of the Dataportability gu ys and am a big supporter of the Privacy and Public Policy work group at Kantara. There is *a lot* of good work being done by folks trying to figure out how to give people greater control over the use of data about them (privacy) and gain access to data they use or created (dataportability).

Unfortunately, sometimes the arguments behind these efforts are based on who owns – or who should own – the data. This is not just an intellectual debate or political rallying call, it often undermines our common efforts to build a better system.

Consider this:

1. Privacy as secrecy is dead
2. Data sharing is data copying
3. Transaction data has dual ownership
4. Yours, mine, & ours: Reality is complicated
5. Taking back ownership is confrontational

Privacy as secrecy is dead

First, the data is pretty much already out there. The issue isn’t “How do we keep data from bad people,” it’s “How do we keep people from doing bad things with data?” DRM and crypto and related technology as the sole means to prevent data leakage and data abuse are failures. Sooner or later, the bad guys break the system and get the data. Sure, there are smart things we can do to protect ourselves. Just like we wear seatbelts and lock our front doors, we should also use SSL and multi-factor authentication, but we can’t count on technology to keep our secrets. We need solutions that work even when the secret is out.

In fact, privacy isn’t about information we keep secret. It is about information we have revealed to someone else with expectation of discretion, e.g., when we tell our doctor about our sexual activities. It’s no longer a secret from the Doctor, but because it is private, we have rules that keep the information from being used inappropriately. Most of the time, with most doctors, it works. Those few who break those rules are dealt with through legal means, both civil and criminal, as well as social approbation. So, because we inherently need to release data to different parties at different times, we can’t control it through secrecy alone. Instead, we need to build a framework for preventing abuse when others *do* have access to sensitive information. Like in the case with our doctor, we want our service providers to have the data they need to provide the highest quality services.

Data sharing is data copying

Second, in the world of atoms, there can only be one of a thing, which is the reverse of the world of bits. With atoms, even if there are copies, each copy is itself a singular thing. Selling, transferring, or stealing a thing precludes the original owner from continuing to use it.

This isn’t true for information, which can easily be sold, transfered, and stolen without disturbing the original version. In fact, the entire Internet is basically a copy machine, copying IP packets from router to router, as we “send” images, web pages, and emails from user to user and machine to machine--each time a new copy is created whether or not the originating copy is deleted. To think of bits as if they were ownable property leads to attempted solutions like DRM that try to technologically prevent access to the information within the data, which is only good until the first hacker cracks the code and distributes it themselves. Instead, if we build social and legal controls on use, we can give information more freely, but under terms set by each individual when they share that information. Enforced by social and legal rather than purely technological means, this makes the most of the low marginal cost of distributing online, while retaining control for contributors.

Transaction data has dual ownership

Third, much interesting data is actually mutually owned… which means the other guy can already do whatever the heck they want with it. Consider web attention data, the stream of digital crumbs representing the websites we’ve visited and any interactions at each: all our purchases, all our blog posts, all our searches. Everything. Some folks argue that we *own* that data and therefore have the right to control the use of it. But so too do the owners of the websites we’ve been visiting. We don’t own our http log entries at Amazon. Amazon does. In fact, in every instance where two parties interact, where we engage in some transaction with someone else, *both* parties are co-creating that information. As such, both parties own it. So, if we tie the issue of control to ownership, then we’ve already lost the battle, because every service provider has solid claims to ownership over the information stored in their log files, just as we, as individuals, own the browsing history stored on our hard drive by Firefox, Internet Explorer and Chrome.

In the movie *Fast Times at Ridgemont High*, in a [confrontation with Mr. Hand|http://slice.seriouseats.com/archives/2010/01/video-jeff-spicoli-classroom-pizza-delivery-in-fast-times-at-ridgemont-high.html], Spicoli argues “If I’m here and you’re here, doesn’t that make it *our* time?” Just like the time shared between Spicoli and Mr. Hand, the information created by visiting a website is co-created and co-owned by both the visitor and the website. Every single interaction between two endpoints on the web generates at least two owners of the underlying data.

This is not a minor issue. The courts have already ruled that if an email is stored for any period of time on a server, the owner of that server has a right to read the email. So, when “my” email is out there at Gmail or AOL or on our company’s servers, know that it is *also*, legally, factually, and functionally, already *their* data.

Yours, mine, & ours: Reality is complicated

Fourth, when two parties come together for any reason, each brings their own data to the exchange. We need a framework that can handle that. Iain Henderson breaks down this complexity in a blog post about your data, my data, and our data, talking about an individual doing business with a vendor, for example, someone buying a car.

“My data” means data that I, as an individual have that is related to the transaction. It could include the kind of car I’m looking for, my budget, and estimates of my spouse’s requirements to approve of a new purchase.

“Your data” means data that the car dealer knows, including the actual cost of the vehicle, the number of units in inventory, the pace of sales, current buzz from other dealers.

“Our Data” means information that both parties have in common. That could be *Shared Information*, explicitly given by one party to the other in the course of the deal, such as a social security number so the dealer could run a credit check. It could be *Mutual Information*, generated by the very act of the transaction, such as the final sale price of the vehicle. Or, it could be *Overlapping Information*, which each party happens to know independently, such as the Manufacturer Suggested Retail Price (MSRP) of a vehicle (which we found online before heading to the dealership).

The ownership of “your” and “my” data is *usually* clear. However, ownership of the different types of “our” data is a challenge at best. To complicate matters further, every instance of “my data” is somebody else’s “your data”. In every case, there is this mutually reciprocal relationship between us and them. In the VRM case, we usually think of the individual as owning “my data” and the vendor as owning “your data”, but for the vendor, the reverse is true: to them their data is “my data” and the individual’s data is “your data”. Similar dynamics occur when the other party is an individual. I bring my data, you bring your data, and together we’ll engage with “our” data. We need an approach that respects and applies to everyone’s data, you, me, them, everybody.

In these complex Venn diagrams of ownership, it is more important who controls the data than who owns it. We’ve already lost the crudest form of control – secrecy – and we are going to continue to lose more as we opt-in to seductive new services based on divulging more and more information: our purchase history, browsing activity, and real-world location data. But we still need to control how all this data is used, to protect our own interests while still enjoying the benefits of the great big copy machine that is the Internet.

Taking back ownership is confrontational

Fifth, we don’t need to pick a fight to change the game. There is a lot of data out there that many of us believe we should have control over. I agree. A lot of people argue that we should have the right to exclude other people’s use because we own the data, because it’s *ours* in some legal, moral, or ethical framework. The problem is, those other people already have it, and they *also* believe that they are legitimate owners. In fact, many of them *paid* for that data, buying it from data aggregators who compile all sorts of things about people, from both public and private sources. This entire ecosystem of customer data is a multi-billion dollar business and every single player “owns” the data they are working with. So if we focus our energy in claiming ownership over that same data in order to take control, we are framing the conversation as a fight, a fight against a powerful, well-healed, well-funded, entrenched bunch of opponents.

Most of these “opponents” are the very people we are trying to win over to our way of thinking. These are the vendors we want to embrace a new way to do business. These are the technologists we want to transform their proven, value-generating CRM systems to work with *our* data on *our* t erms, instead of *their* data on *their* terms. Arguing over ownership puts these potential allies on the defensive, when what we really want is their collaboration.

From Ownership to Authority, Rights, and Responsibilities

Rather than building a regime based on data ownership, I believe we would be better served by building one based on authority, rights, and responsibilities. That is, based on Information Sharing.

- Who has the authority to control access and use of particular information?
- What rights does a party have in using and distributing a piece of information?
- What responsibilities does an information user have to others with respect to that information?

Let’s stop arguing about who owns what and start figuring out how we can share information in ways that allow everyone to win.

When we collect all of our information into a single conceptual repository, and then share access to it with service providers on our own terms, we create a high quality, highly relevant, curated personal datastore. This allows us to bootstrap a control regime over all of our data in a way that creates new value for us and for our service providers. Now, instead of iTunes Genius or a Last.FM scrobbler only having access to our media use with their service, they can provide recommendations based on all the information stored in our personal audio datastore. We get better recommendations and they get better data to drive their services. This personal datastore is entirely under the authority of the user, sharing information with service providers according to specific rights and responsibilities.

The Information Sharing approach neatly sidesteps the complexities involved in privacy and dataportability issues of the information already known by service providers. These remain serious issues, worth addressing. Resolving them will require long term investment in the legal, regulatory, moral, and political systems that govern our society. Fortunately, sharing the information in our personal datastore can begin almost immediately once we have working specifications.

This controlled sharing of information will dramatically increase our comfort level when revealing our intentions and interests. We would have control over the use – and would be able to prevent abuse – of that information, while making it easy for service providers to improve our lives in countless ways.

At the Information Sharing Work Group at the Kantara Initiative, Iain Henderson and I are leading a conversation to create a framework for sharing information with service providers, online and off. We are coordinating with folks involved in privacy and dataportability and distinguish our effort by focusing on new information, information created for the purposes of sharing with others to enable a better service experience. Our goal is to create the technical and legal framework for Information Sharing that both protects the individual and enables new services built on previously unshared and unsharable information. In short, we are setting aside the questions of data ownership and focusing on the means for individuals to control that magical, digital pixie dust we sprinkle across every website we visit.

Because the fact is, we *want* to share information. We want Google to know what we are searching for. We want Orbitz to know where we want to fly. We want Cars.com to know the kind of car we are looking for.

We just don’t want that information to be abused. We don’t want to be spammed, telemarketed, and adverblasted to death. We don’t want companies stockpiling vast data warehouses of personal information outside of our control. We don’t want to be exploited by corporations leveraging asymmetric power to force us to divulge and relinquish control over our addresses, dates of birth, and the names of our friends and family.

What we want is to share our information, *on our terms*. We want to protect our interests *and* enable service providers to do truly amazing things for us and on our behalf. This is the promise of the digital age: fabulous new services, under the guidance and control of each of us, individually.

And that is precisely what Information Sharing work group at Kantara is enabling.

The work is a continuation of several years of collaboration with Doc Searls and others at ProjectVRM. We’re building on the principles and conversations of Vendor Relationship Management and User Driven Services to create an industry standard for a legal and technical solution to individually-driven Information Sharing.

Our work group, like all Kantara work groups, is open to all contributors – and non-contributing participants – at no cost. I invite everyone interested in helping create a user-driven world to join us.

It should be an exciting future.

This article first published independently by Joe Andrieu at Beyond Data Ownership to Information Sharing on January 21, 2010. Submitted to the Kantara Information Sharing Work Group for consideration and further development. on March 1, 2010.