The cookie turned 20 just the other day. More than a tasty morsel of technology, two decades of experimentation have created considerable value around its use.

The cookie originated with the ninth employee of Netscape, Lou Montulli. Fresh out of college in June 1994, Montulli sought to embed a user’s history in a browser’s functions. He added a simple tool, keeping track of the locations users visited. He called his tool a “cookie” to relate it to an earlier era of computing, when systems would exchange data back and forth in what programmers would call “magic cookies.” Every browser maker has included cookies ever since.

The cookie had an obvious virtue over many alternatives: It saved users time, and provided functionality that helped complete online transactions with greater ease. All these years later, very few users delete them (to the disappointment of many privacy experts), even in the browsers designed to make it easy to do so.

Montulli’s invention baked into the Web many questions that show up in online advertising, music, and location-based services. Generating new uses for information requires cooperation between many participants, and that should not be taken for granted.

The cookie’s evolution

Although cookies had been designed to let one firm track one user at a time, in the 1990s many different firms experimented with coordinating across websites in order to develop profiles of users. Tracking users across multiple sites held promise; it let somebody aggregate insights and achieve a survey of a user’s preferences. Knowing a user’s preferences held the promise of more effective targeting of ads and sales opportunities.

DoubleClick was among the first firms to make major headway into such targeting based on observation at multiple websites. Yet, even its efforts faced difficult challenges. For quite a few years nobody ever targeted users with any precision, and overpromises fueled the first half-decade of experiments.

The implementation of pay-per-click and the invention of the keyword auction—located next to an effective search engine—brought about the next great jump in precision. That, too, took a while to ripen, and, as is well known, Google largely figured out the system after the turn of the millennium.

Today we are awash in firms involved in the value chain to sell advertising against keyword auctions. Scores stir the soup at any one time, some using data from cookies and some using a lot more than just that. Firms track a user’s IP addresses, and the user’s Mac address, and some add additional information from outside sources. Increasingly, the ads know about the smartphone’s longitude and latitude, as well as an enormous amount about a user’s history.

All the information goes into instantaneous statistical programs that would make any analyst at the National Security Agency salivate. The common process today calculates how alike one individual is to another, assesses whether the latest action alters the probability the user will respond to a type of ad, and makes a prediction about the next action.

Let’s not overstate things. Humans are not mechanical. Although it is possible to know plenty about a household’s history of surfing, such data can make general predictions about broad categories of users, at best. The most sophisticated statistical software cannot accurately predict much about a specific household’s online purchase, such as the size of expenditure, its timing, or the branding.

Online ads also are still pretty crude. Recently I went online and bought flowers for my wedding anniversary and forgot to turn off the cookies. Not an hour later, a bunch of ads for flowers turned up in every online session. Not only were those ads too late to matter, but they flashed later in the evening after my wife returned home and began to browse, ruining what was left of the romantic surprise.

Awash in metadata

Viewed at a systemic level, the cookie plays a role in a long chain of operations. Online ads are just one use in a sizable data-brokerage industry. It also shapes plenty of the marketing emails a typical user receives, as well as plenty of offline activities, too.

To see how unique that is, contrast today’s situation with the not-so-distant past.
Consider landline telephone systems. Metadata arises as a byproduct of executing normal business processes. Telephone companies needed the information for billing purposes—for example, the start and stop time for a call, area codes and prefix to indicate originating and ending destination, and so on. It has limited value outside of the stated purpose to just about everyone except, perhaps, the police and the NSA.

Now contrast with a value chain involving more than one firm, again from communications, specifically, cellular phones. Cell phone calls also generate a lot of information for their operations. The first generation of cell phones had to triangulate between multiple towers to hand off a call, and that process required the towers to generate a lot of information about the caller’s location, the time of the call, and so on.

Today’s smartphones do better, providing the user’s longitude and latitude. Many users enable their smartphone’s GPS because a little moving dot on an electronic map can be very handy in an unfamiliar location (for example). That is far from the only use for GPS.

Cellular metadata has acquired many secondary values, and achieving that value involves coordination of many firms, albeit not yet at an instantaneous scale suggestive of Internet ad auctions. For example, cell phone data provides information about the flow of traffic in specific locations. Navteq, which is owned by the part of Nokia not purchased by Microsoft, is one of many firms that make a business from collecting that data. The data provide logistics companies with predictable traffic patterns for their planning.

Think of the modern situation this way: One purpose motivated collecting metadata, and another motivated repurposing the metadata. The open problem focuses on how to create value by using the data for something other than its primary purpose.

**Metadata as a source of value**

Try one more contrast. Consider a situation without a happy ending.

New technologies have created new metadata in music, and at multiple firms. Important information comes from any number of commercial participants—ratings sites, online ticket sales, Twitter feeds, social networks, YouTube plays, Spotify requests, and Pandora playlists, not to mention iTunes sales, label sales, and radio play, to name a few.

The music market faces the modern problem. This metadata has created a great opportunity. The data has enormous value to a band manager making choices in real time, for example. Yet, the entire industry has not gotten together to coordinate use of metadata, or even to coordinate on standard reporting norms.

There are several explanations for the chaos. Some observers want to blame Apple, as it has been very deliberate about which metadata from iTunes it shares, and which it does not. However, that is unfair to Apple. First, they are not entirely closed, and some iTunes data does make it into general use. Moreover, Apple does not seem far out of step with industry practices for protecting one’s own self-interest, which points to the underlying issue, I think.

There is a long history of many well-meaning efforts being derailed by narrow-minded selfishness. For decades, merely sampling another performer’s song in any significant length led to a seemingly trivial copyright violation that should have been easy to resolve. Instead, the industry has moved to a poor default solution, requiring samplers to give up a quarter of royalties. With those type of practices, there is very little sampling. That seems suboptimal for a creative industry.

Composers and performers also have had tussles for control over royalties for decades, and some historical blowups took on bitter proportions. The system for sharing royalties in the US today is not some grand arrangement in which all parties diplomatically compromised to achieve the greater good. Rather, the system was put there as a consent decree after settling an antitrust suit.

If this industry had a history of not sharing before the Internet, who thought the main participants would share metadata? Who would have expected the participants to agree on how to aggregate those distinct data flows into something useful and valuable? Only the most naive analyst would expect a well-functioning system to ever emerge out of an industry with this history of squabbling.

More generally, any situation involving more than a few participants is ripe for coordination issues, conflict, and missed opportunity. It can be breathtaking when cooperation emerges, as in the online advertising value chain. That is not a foregone conclusion. Some markets will fall into the category of “deals waiting to be done.”

The systems are complicated, but the message is simple. Twenty years after the birth of the cookie, we see models for how to generate value from metadata, as well as how not to. Value chains can emerge, but should not be taken for granted.

More to the point, many opportunities still exist to whip up a recipe for making value from the new data layer, if only the value chain gets organized. On occasion, that goal lends itself to the efforts of a well-managed firm or public efforts, but it can just as easily get neglected by a squabbling set of entrepreneurs and independently minded organizations, acting like too many cooks.

**Reference**


Shane Greenstein is the Kellogg Chair in Information Technology at the Kellogg School of Management, Northwestern University. Contact him at greenstein@kellogg.northwestern.edu.