

The War Over the Value of Personal Data

In a world increasingly dependent on turning personal data into profits, it is unclear how much that data is actually worth.

DATA HAS BEEN called the “new oil,” and one reason for this is that personal data greases the wheels of our connected world. It powers wildly lucrative social media platforms like Facebook and Snapchat. Data makes online advertising super-targeted (and super-profitable) for Internet giants like Google, and data about online habits is highly lucrative for any brand selling online (which is most of them).

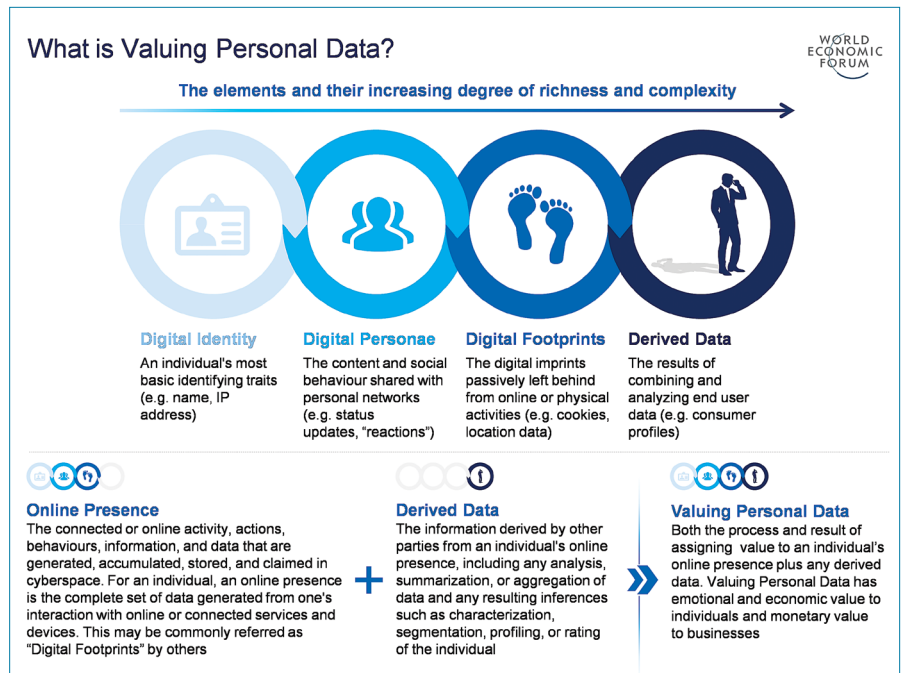
If data is the new oil, we’re discovering gushers each and every day. Consultancy IDC predicts the total amount of data generated globally will hit 44 zettabytes by 2020, a tenfold jump from 2013’s 4.4 zettabytes.

The value of this new oil has been enhanced by artificial intelligence (AI) and machine learning systems that are able to make sense of it all. As it turns out, machines are better at extracting value from structured and unstructured data than humans.

No matter how adept humans or machines are at giving data value, how much our data is actually worth remains an open question. In theory, data is everywhere, available to any company with the infrastructure to leverage it for commercial gain. The reality, however, is a little messier.

Individuals have uncertain control over how their data is collected, viewed, and monetized. Companies that want to earn profits from data, through AI or traditional data analysis, also face some obstacles.

“Companies, especially in industries such as financial services and health-care, have significant barriers to monetizing data, such as confidentiality, privacy, and regulatory requirements,” says Colton Jang, cofounder of LeapYear Technologies, a company that develops technology that enables firms to analyze and monetize sensitive data legally.



The result? A war is under way over data, but it’s not entirely clear how much the resource is actually worth.

Quid Pro Quo

What is clear is that the volumes of data being generated are increasing. Cisco predicts a growth in annual Internet traffic of 175% over the next half-decade, from 1.2 zettabytes a year in 2016 to 3.3 zettabytes in 2021. While this data explosion is significant, it isn’t clear how much of the data is “useful or valuable,” reports *Semiconductor Engineering*.

“A lot of industries have figured out that their business, product, and business models could be impacted by a different utilization of the data that is somehow attached to their devices or their business models,” Aart de Geus, chairman and co-CEO of software company Synopsys, told *Semiconductor Engineering*.

“If you can harness that in a way that finds shortcuts and efficiencies, or

just completely different ways of going about business, that is high-impact.”

This has firms in nearly every industry looking for ways to both acquire, and extract value, from consumer data. One of the major fronts in this is the home.

IDC predicts the number of devices connected to the Internet will triple by 2020 to 30 billion, and nearly triple again five years later. Many of these “Internet of Things” (IoT) devices are smart home appliances and tools, like connected refrigerators or general-purpose voice assistants like Amazon’s Echo device, which is powered by the company’s Alexa machine learning system.

IoT devices rely on data to drive user adoption. Your smart fridge may start with a series of assumptions about your buying habits based on past purchase data, then guess at what items to automatically reorder. To keep you happy, the makers of the fridge must continually ingest data on your habits

ACM Journal of Data and Information Quality



Providing Research and Tools
for Better Data

ACM JDIQ is a multi-disciplinary journal that attracts papers ranging from theoretical research to algorithmic solutions to empirical research to experiential evaluations. Its mission is to publish high impact articles contributing to the field of data and information quality (IQ).



For further information
or to submit your
manuscript,
visit jdiq.acm.org

to improve the quality of their guesses about what you want to buy next.

If they succeed, you keep using the fridge. If they do not, it's off to a competitor.

In all cases, data is the oil that powers the device. It is the resource that enables the device to improve—and thus attract more customers. This makes your personal data extremely valuable.

The question is: How valuable?

Consumers are potentially willing to give up their data—but they want something in return. For instance, a study conducted by Parks Associates found approximately half of U.S. households are willing to share smart device data and control in their home in exchange for a discount on electricity.

However, not every price is one consumers are willing to pay. The study also found that consumers were more likely to share their data in exchange for incentives like discounts, versus “intangibles” such as “product recommendations or simplified ordering.”

This seems like a pretty straightforward quid pro quo: companies need to make it worthwhile for consumers to relinquish their personal data. In the case of smart homes, maybe that's a discount on bills. For a social network like Facebook or an Internet giant like Google, users are given state-of-the-art communication and search tools in exchange for data on their browsing habits.

In some cases, companies and consumers agree on the value of data and transact in kind. But even when they do, turning consumer data into dollars can be difficult. The real value of data may actually lie in its aggregation. Tech titans are looking to learn about millions of people at once, not individuals, and they value data that has been analyzed in aggregate to deliver insights that can be monetized, rather than a muddle of machine-generated data that hasn't been assessed.

This leads to wildly different assessments of the monetary value data provides. Business analytics student Pauline Glikman and econometrics professor Nicolas Gladly tried to assess data's value in a 2015 *TechCrunch* article. In assessing Facebook acquisitions of WhatsApp and Instagram, as well as Microsoft's purchase of Minecraft, they found the value of each user was anywhere between \$15 and \$40. How-

ever, general information about individuals, like age or gender, was sold for as little as \$0.0007 per data point by data brokers that collected this information online.

So how much is data actually worth? The short answer is, it depends. This uncertainty introduces some complexities when governments attempt to introduce blanket data protection regulations.

Regulatory Hurdles

In the European Union, the General Data Protection Regulation (GDPR) will come into force in May. In the opinion of Jang at LeapYear Technologies, this law is the most significant one affecting individual and corporate data. “Any company that collects or processes data on EU citizens will need to comply with strict new requirements for data protection, or face massive fines for non-compliance,” he says.

The GDPR protects any information “that can be used to directly or indirectly identify the person,” according to the European Union's official website on the regulation. This includes “anything from a name, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer IP address.”

The rights conferred by the GDPR to EU citizenry include the right to be notified of data breaches, visibility into how their personal data is used by companies that collect it, and the “right to be forgotten,” or the right to request that whom-ever has their personal data erase it.

This law gives individuals increased control over their personal data and, by extension, its value to firms who want to access it. Companies that sell to EU citizens will no longer be sure the data they collect will be available indefinitely. In fact, says Jang, companies that process large volumes of sensitive data, use data to predict or profile, and/or transfer sensitive data across borders should be preparing now.

“Companies will need to incorporate anonymization, data minimization, and privacy-by-design into their data processing activities,” Jang says. The GDPR can apply to companies outside the EU, which puts a large swath of firms at risk. In fact, any company that offers goods and services to EU citizens must comply with the new regulatory environment.

An Uneasy Peace

There may not need to be a war over the value of personal data; or, at least, not one that pits consumers directly against the firms to which they give their money.

Companies like LeapYear Technologies are helping to navigate the gap between the needs of companies and the regulations designed to protect individuals. LeapYear develops cryptographic technology that enables statistical analysis of a dataset without disclosing information about individual records. “Analysts can use our API to compute reports, statistics, and machine learning models against the data without being able to view or extract any information from the underlying data source,” Jang says.

This kind of compromise may be necessary. For one thing, it’s extremely difficult for companies in the connected era to comply fully with regulations like GDPR.

“Almost all data monetization strategies involve repurposing existing data assets that were originally collected for another purpose,” says Jang. “Under GDPR, this cannot be done without explicit opt-in consent from each citizen.”

On the other hand, it is easier than ever to collect data on a user’s online habits from anywhere in the world—in real time.

With cryptographically anonymized datasets like the ones LeapYear produces, consumers could enjoy the benefits of free online platforms and services without worrying about their data being abused. Brands could monetize that data and improve products with confidence, certain they won’t incur huge downside risk by doing so.

After all, the value of personal data is uncertain for companies precisely because the cost of noncompliance is oh-so-dear.

“Processing personal data in a non-compliant manner [under the GDPR] can result in significant fines, the greater of 20,000,000 Euros or 4% of global revenues,” says Jang. “Given the magnitude of the penalty, companies that collect or process data in the EU will need to reduce the scope of their data strategy or overhaul core business processes to incorporate new privacy-enhancing technology.”

On paper, consumer data is valuable

Companies that collect or process data in the EU will need to reduce the scope of their data strategy or incorporate privacy-enhancing technology into core business processes.

to firms that require it for better products and competitive advantages. In reality, much like the first rush of wildcaters and opportunists looking to find oil in unlikely places, Internet firms are finding the penalty for drilling in the wrong spot is costly indeed. **Q**

Further Reading

Baar, A.

Smart Home Consumers Willing To Give In Order To Get

MediaPost, July 26, 2017

<http://bit.ly/2w2NSE9>

Glikman, P., and Gladly, N.

What’s the Value of Your Data?

TechCrunch, October 13, 2015

<http://tcrn.ch/2xBiBux>

Kanellos, M.

152,000 Smart Devices Every Minute In 2025: IDC Outlines The Future of Smart Things,

Forbes, March 13, 2016

<http://bit.ly/2vnXGXg>

Li, C., Li, D.Y., Miklau, G., and Suci, D.

A Theory of Pricing Private Data

Communications, December 2017,

<http://bit.ly/2j57NMU>

Sperling, E.

The Rising Value of Data

Semiconductor Engineering, August 8, 2017

<http://bit.ly/2vf9qMz>

EU General Data Protection Regulation Portal

<http://www.eugdpr.org/>

Logan Kugler is a freelance technology writer based in Tampa, FL, USA. He has written for over 60 major publications.

© 2018 ACM 0001-0782/18/2 \$15.00

ACM Member News

TRACKING DOWN THE PHYSICAL INTERNET



“I have always been interested in math and puzzles,” says Paul Barford, professor of Computer

Science at the University of Wisconsin-Madison. “The path that led me to computer science was based on recognizing the mathematical foundation of computer science.”

Barford earned his B.S. in electrical engineering from the University of Illinois at Urbana-Champaign, then worked in industry for eight years before completing his Ph.D., in computer science at Boston University.

His research interests focus on collecting data and then applying analytic techniques to it. “I enjoy this very practical aspect of research,” Barford says. “I like to get my hands on equipment, and data.”

For the past six years, Barford has been working on the Internet Atlas Project, a comprehensive repository of the physical Internet being developed by the Center for Applied Internet Data Analysis. “I am trying to figure out where the Internet is actually deployed—the physical conduits—the places where those conduits are actually connected to communication devices,” he explains.

A colleague described his Internet Atlas work as “Internet archeology.” Barford says it was a lot of work to develop techniques to identify the physical paths of Internet data, as well as the nodes at which they terminate. His work on Internet topologies has allowed him to apply this data to understand how the Internet can be made more robust and secure.

“I am interested in how users are actually interacting with the Internet,” Paul adds. “At the end of the day, we are creating technology to serve users.”

—John Delaney