If the last century was marked by the ability to observe the interactions of physical matter—think of technologies like x-ray and radar—this century, he says, is going to be defined by the ability to observe people through the data they share.

So-called anticipatory systems such as Google Now represent one example of what could result. We’re already seeing the transformations that big data is causing in advertising and other situations where millions of people’s activity can be measured at a time. Now data science is looking at how it can help individuals. Timely updates on a United Airways flight may be among the tamer applications. Think instead of statistical models that tell you what job to take, or alert you even before you feel ill that you may have the flu.

Driving this trend is a swelling amount of personal data available to computers. The amount of digital data being created globally is doubling every two years, and the majority of it is generated by consumers, in the form of movie downloads, VoIP calls, e-mails, cell-phone location readings, and so on, according to the consultancy IDC. Yet only about 0.5 percent of that data is ever analyzed.

“There is so much more data out there that you can afford to tailor it to the individual,” says Patrick Wolfe, a statistician who studies social networks at University College, London. “Statistically, strength comes from pooling people together, but then the icing on the cake is when you individualize the findings.”

For the data refineries of Silicon Valley, like Google, Facebook, and LinkedIn, the merger of big data and personal data has been a goal for some time. It creates tools advertisers can use, and it makes products that are particularly “sticky,” too. After all, what’s more interesting than yourself? Facebook suggests who your friends might be. Google Now gets better the more data you give it.

Exposing more personal data seems inevitable. With the huge jump in sales of smartphones packed with accelerometers, cameras, and GPS, “people have become instrumented to collect and transmit personal data,” says Weigend. And that may be just the start. Already a fringe community of technophiles in what’s known as the quantified-self movement have been equipping their bodies with sensors, pedometers, even implanted glucose monitors.

One technophile we feature in this report is Stephen Wolfram, the creator of the search engine Wolfram Alpha, who has for years engaged in a massive self-tracking project, cataloguing e-mails, keystrokes, even his physical movements. Wolfram is interested in predictive apps but also in the insights that large data sets can have on personal behavior, something he calls “personal analytics.” Wolfram’s idea is that just as his search engine tries to organize all facts about the world, “what you have to do in personal analytics is try to accumulate the knowledge of a person’s life.”

The holdup, says Wolfram, is that some of the most useful data isn’t being captured, at least not in a way that’s easily accessible. Part of the problem is technical, a lack of integration. But much data is warehoused by private companies like Facebook, Apple, and Fitbit, maker of a popular pedometer. Now, as the value of personal data becomes more apparent, fights are brewing. California legislators this year introduced a “Right to Know” bill that would require companies to reveal to individuals the “personal information” they store—in other words, a digital copy of every location trace and sighting of their IP address.

The bill is a part of a social movement that is demanding privacy and accountability but also a different economic arrangement between the people who supply the data and those who apply it. People want more of the direct benefits of big data, and this month’s MIT Technology Review Business Report tracks the technology, apps, and business ideas with which industry is responding.

—Antonio Regalado

### Emerged Technologies

### Has Big Data Made Anonymity Impossible?

As the amount of data expands exponentially, nearly all of it carries someone’s digital fingerprints.

- In 1995, the European Union introduced privacy legislation that defined “personal data” as any information that could identify a person, directly or indirectly. The legislators were apparently thinking of things like documents with an identification number, and they wanted them protected just as if they carried your name.

Today, that definition encompasses far more information than those European legislators could have imagined—more than all the bits and bytes in the entire world when they wrote their law 18 years ago.

Here’s what happened. First, the amount of data created each year has grown exponentially: it reached 2.8 zettabytes in 2012, a number that’s as gigantic as it sounds, and will double again by 2015, according to the consultancy IDC. Of that, about three-quarters is generated by individuals as they create and move digital files. A typical American office worker produces 1.8 million megabytes of data each year. That is about 5,000 megabytes a day, including downloaded movies, Word files, e-mail, and the bits generated by computers as that information is moved across mobile networks or the Internet.

Much of this data is invisible to people and seems impersonal. But it’s not. What modern data science is finding is that nearly any type of data can be used, much like a fingerprint, to identify the person who created it: your choice of movies on Netflix, the location signals emitted by your cell phone, even your pattern of walking as recorded by a surveillance camera. In effect, the more
data there is, the less any of it can be said to be private, since the richness of that data makes pinpointing people "algorithmically possible," says Princeton University computer scientist Arvind Narayanan.

We're well down this path already. The information we thought of as personal data in the past—our name, address, or credit card records—is already bought and sold by data brokers like Axiom, a company that holds an average of 1,500 pieces of information on more than 500 million consumers. This is data that people put into the public domain on a survey form or when they signed up for services such as TiVo.

Axiom uses information about the make and year of your car, your income and investments, and your age, education, and zip code to place you in one of 70 different "PersonieX" clusters, which are "summarized indicators of lifestyle, interests, and activities." Did you just finalize a divorce or become an empty nester? Such "life events," which move people from one consumer class to another, are of key interest to Axiom and its advertising clients. The company says it can analyze its data to predict 3,000 different propensities, such as how a person may respond to one brand over another.

Yet these data brokers today are considered somewhat old-fashioned compared with Internet companies like Facebook, which have automated the collection of personal information so it can be done in real time. According to its financial filings at the time of its IPO, Facebook stores around 111 megabytes of photos and videos for each of its users, who now number more than a billion. That's 100 petabytes of personal information right there. In some European legal cases, plaintiffs have learned that Facebook's records of their interactions with the site—including text messages, things they "liked," and addresses of computers they used—run to 800 printed pages, adding up to another few megabytes per user.

In a step that's worrisome to digital-privacy advocates, offline and online data sets are now being connected to help marketers target advertisements more precisely. In February, Facebook announced a deal with Axiom and other data brokers to merge their data, linking real-world activities to those on the Web. At a March investor meeting, Axiom's chief science officer claimed that its data could now be linked to 90 percent of U.S. social profiles.

Such data sets are often portrayed as having been "anonymized" in some way, but the more data they involve, the less likely that is to be actually true. Mobile-phone companies, for instance, record users' loca-

---

**65 billion**
Location-tagged payments made in the U.S. annually

**154 billion**

E-mails sent per day

**87%**

U.S. adults whose location is known via their mobile phone

---

**Digital Information Created Each Year, Globally**

2,000 BILLION GIGABYTES

<table>
<thead>
<tr>
<th>Year</th>
<th>Data (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>400</td>
</tr>
<tr>
<td>2006</td>
<td>600</td>
</tr>
<tr>
<td>2007</td>
<td>800</td>
</tr>
<tr>
<td>2008</td>
<td>1,000</td>
</tr>
<tr>
<td>2009</td>
<td>1,200</td>
</tr>
<tr>
<td>2010</td>
<td>1,400</td>
</tr>
<tr>
<td>2011</td>
<td>1,600</td>
</tr>
</tbody>
</table>

**2,000%**
Expected increase in global data by 2020

---

**III Megabytes**
Video and photos stored by Facebook, per user

---

**75%**
Percentage of all digital data created by consumers
tions, strip out the phone numbers, and sell aggregate data sets to merchants or others interested in people's movements. MIT researchers Yves-Alexandre de Montjoye and César A. Hidalgo have shown that even when such location data is anonymous, just four different data points about a phone's position can usually link the phone to a unique person.

The greater the amount of personal data that becomes available, the more informative the data gets. In fact, with enough data, it's even possible to discover information about a person's future. Last year Adam Sadilek, a University of Rochester researcher, and John Krumm, an engineer at Microsoft's research lab, showed they could predict a person's approximate location up to 80 weeks into the future, at an accuracy of above 80 percent. To get there, the pair mined what they described as a "massive data set" collecting 32,000 days of GPS readings taken from 307 people and 396 vehicles.

Then they imagined the commercial applications, like ads that say "Need a haircut? In four days, you will be within 100 meters of a salon that will have a $5 special at that time." Sadilek and Krumm called their system "Far Out." That's a pretty good description of where personal data is taking us. —Patrick Tucker

Emerged Technologies

Predictive Apps Mine Your Life

In a break from traditional software, new apps offer information proactively.

- A new type of mobile app is departing from a long-standing practice in computing. Typically, computers have just dumbly waited for their human operators to ask for help. But now applications based on machine learning software can speak up with timely information even without being directly asked for it. They might automatically pull up a boarding pass for your flight just as you arrive at the airport, or tell you that current traffic conditions require you to leave for your next meeting within 10 minutes.

The highest-profile of these apps is Google Now, which is a feature of the latest version of the Android mobile operating system and was recently added to the Google search app for the iPhone. Google Now is trained to predict when a person is about to take certain actions and offer help accordingly. It can also learn about an individual to fine-tune the assistance it offers.

Google Now's algorithms use the data in a person's Google e-mail and calendar accounts and Web searches. The app learns where you live and work and when you commute so that it can offer a virtual index card showing traffic or transit information. Other cards offer boarding passes and other handy information at appropriate times.

Bill Ferrell, founder and CEO of Osito, a company with an iPhone app that offers similar functions, calls this idea "predictive intelligence." Osito's system foretells a person's actions and needs from location, e-mail, and calendar data and uses those predictions to go beyond offering just advice. It also presents ways for a person to take action. A flight reminder will include a button to summon a cab, for example.

Now that the first generation of this type of app has been well received, engineers at Google, Osito, and elsewhere seek to wring more insights from the data they collect about their users. Osito's engineers are working to learn more from a person's past location traces to refine predictions of future activity, says Ferrell. Google Now recently began showing the weather in places it believes you're headed to soon. It can also notify you of nearby properties for

<table>
<thead>
<tr>
<th>NAME</th>
<th>Cue</th>
<th>Google Now</th>
<th>Osito</th>
<th>Tempo AI</th>
<th>Dark Sky</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAISED</td>
<td>$4.7 million</td>
<td>N/a</td>
<td>$1.1 million</td>
<td>Incubated at SRI International</td>
<td>$39,376</td>
</tr>
<tr>
<td>PREDICTIONS</td>
<td>Summarizes a person's day based on information scavenged from calendar, e-mail, and documents</td>
<td>Directions, traffic, and weather based on a person's location and calendar</td>
<td>Handles transactions like checking in for a flight or calling a cab after you land at the airport</td>
<td>Directions to appointments. Also sends messages if you're running late</td>
<td>Provides minute-by-minute weather forecasts for user's exact location</td>
</tr>
</tbody>
</table>