

BY ERICA ORANGE



WFS PHOTO ILLUSTRATION BY C. G. WAGNER / PHOTOS.COM

# MINING INFORMATION from the Data Clouds

THIS CLOUD OF DATA THAT WE DAILY CONTRIBUTE TO MAY YIELD A WEALTH OF NEW, VITAL INFORMATION. "CLOUD MINING" MAY SOON ALLOW US TO PREDICT BEHAVIORS OF THE MASSES AND EVEN OFFER ADVICE, ACCORDING TO A BUSINESS FUTURIST.

While it has never been easy to predict human behavior, it is becoming easier, thanks to a variety of technological advances. Your new favorite gadgets are recording much more about you than you may think: your daily habits, your social network, aspects of your personality, and even who you meet and where you go.

The Internet, as it moves to consolidate all forms of media, is providing an incredibly detailed register of our lives. We all leave a trail of digital bread crumbs behind us. These crumbs are easily collected, can be stored indefinitely, and can be repro-

duced and distributed with ease, almost without cost.

Services such as Whrrl, Buddycloud, Brightkite, and Loopt are all examples of location-aware mobile social networks that enable users to see different information based on either where they or where their networks are located. These applications enable mobile phones to combine social networking and instant messaging with the ability to precisely pinpoint and show their own location on an interactive map. The latter process, known as geotagging, can be done on almost any GPS-enabled de-

vice. Programs designed to record the locations that mobile phone users visit also allow the wireless industry, among others, to learn more about the daily behavior of both consumers and their networks.

The number of mobile phones will soon exceed 3.3 billion. Globally, mobile phone penetration is expected to reach 75% by 2011, and the mobile is on course to replace the PC as the primary device for getting online. Every time we use our mobile phones or any other GPS-enabled device, our location is picked up by nearby towers. Since people rou-

## TYPES OF CLOUDS

- **The Public Cloud:** what most likely springs to mind when you think of the “cloud.” Businesses and consultants often use the term “cloud” more or less interchangeably with the Internet. More specifically, however, it refers to an open, externally located, publicly accessible cloud-computing environment whereby online third-party providers either rent out computing resources, services, and applications (usually on demand or on a subscription basis) or make them freely available to businesses and individuals.

- **The Private Cloud:** a custom-designed cloud computing-like environment within a protected firewall on a closed internal network. In this case, the cloud architecture is purchased, rather than rented. Boasting tighter security as well as greater reliability, private clouds are quite costly, and thus not an option for most start-ups and businesses, who will continue to rely on the public cloud. Critics claim that “private cloud” is a contradiction in terms, arguing that cloud computing by definition relies on external providers in an open computing environment.

- **The Hybrid Cloud:** a combination of private/internal and public/external resources, services, and applications that can boost a company’s existing computer infrastructure and at least claim to offer the best of both worlds. Hybrid clouds may ultimately become the industry norm.

—Aaron M. Cohen

tinely carry their mobile phones with them, the device’s very nature makes it an ideal tool for studying both individuals and organizations.

These devices are now shedding light not only on how information is transmitted back and forth but also on how social networks evolve. Patterns in the data can be identified and then translated into maps of social relationships. Location-based technology is becoming so sophisticated that it can predict not only when someone in your network is close by, but also where and when you might like to meet up with that certain someone. One day soon, your iPhone may even be able to play matchmaker, recommending that you introduce yourself to a nearby stranger with whom you have much in common. (That application, Serendipity, is currently in the process of being commercialized by a group of MIT researchers.)

## MINING FOR REALITY

Social scientists have long struggled to create comprehensive and predictive models of human social dynamics. The speedy and widespread adoption of mobile phone technology allows social scientists to collect a much larger, and much more unbiased, assortment of data on human behavior. The collection of real-time behavioral data adds a new dimension. As human interaction becomes increasingly virtual, our ability to analyze speech, interpret movements, and anticipate behavior will compound exponentially.

Social networking sites such as Facebook and MySpace, lifecasting services such as Twitter, social bookmarking services such as Delicious, and digital media sites such as YouTube are all creating a dense jungle of information that is nearly impossible to sift through. While each small particle of information sent into the cloud may seem insignificant on its own, over time, these individual pieces of information coalesce into highly sophisticated portraits of peoples’ lives.

“Reality mining,” a term coined by MIT professor Sandy Pentland, author of *Honest Signals: How They Shape Our World* (The MIT Press,

2008), plays on the term “data mining” and refers to the collection and analysis of technology-based data as it relates to human social behavior. This information is collected mainly from sensors embedded in cell phones. Researchers are working to create algorithms that can identify patterns in the data and translate that information in ways that allow us to learn more about individual and group behavior on both a micro- and a macroscale.

Understanding how clusters of people behave has useful implications not only in terms of, say, analyzing traffic patterns during rush hour, but also for such fields as criminal intelligence, community formation, and disease control. For example, Google’s philanthropic unit, google.org, has introduced a Web tool called Google Flu Trends, which is based on the idea that people feeling sick are likely to turn to the Internet for information. The system is able to track queries entered into the search engine, thereby ultimately predicting regional outbreaks of the flu. Reality-mining methods may also be used to predict unexpected conflicts and emerging problems in the workplace and to analyze employee satisfaction, productivity, and financial risks. A Web service called Pachube enables people around the world (and their virtual-world avatars as well) to share real-time sensor data, such as the amount of energy being consumed in someone’s home, or wind speed and climate in a remote area, via tiny wireless sensors.

The demand for services built around location is growing. Now that we are able to pinpoint and evaluate location information accurately and in real time, figuring out how to leverage this technology as a product or service will lead to a distinct competitive advantage. We will increasingly be able to track both macro- and microtrends as they relate to everything from consumer spending habits to criminal behavior. Since even the smallest of changes in the environment can influence behavior, mobile phones may eventually monitor users’ physical activity and the rhythms of their voice. Embedded technologies (such as speech-analysis software and motion sen-



## FOR FURTHER READING

*Dot Cloud: The 21st Century Business Platform Built on Cloud Computing* by Peter Fingar, Meghan-Kiffer Press, 2009.

*Honest Signals: How They Shape Our World* by Sandy Pentland, The MIT Press, 2008.

*Inescapable Data: Harnessing the Power of Convergence* by Chris Stakutis and John S. Webster, IBM Press, 2005.

*Programming Collective Intelligence: Building Smart Web 2.0 Applications* by Toby Segaran, O'Reilly Media, 2007.

**LEFT:** A rack of commodity computers, part of a data center. The Sector System, a cloud computing system designed by the National Center for Data Mining, links remote data centers together directly, which is faster and cheaper than having them feed into a central location.

**BOTTOM:** A scene from Dreamforce 2008, Salesforce.com's sixth annual user and developer conference in San Francisco, California. The conference bills itself as "The Cloud Computing Event of the Year."



sors) may even detect early warning signs of debilitating ailments, like Parkinson's disease.

Reality mining also enables us to gain unique insight into the collaborative and communicative styles within an organization. These methods may increasingly allow organizations to observe the gradual evolution of relationships, and allow management and departmental heads to analyze the dynamics of networks that could not previously be observed. For example, reality mining could be utilized to determine how well people work within groups and how well employees get along (or don't get along) with one another. Employers could keep track of how many times an employee visits the water cooler or takes a break to chat with co-workers. It could also be used to measure cross-departmental and cross-functional interaction and communication.

## LIVING INSIDE THE CLOUD

Reality mining doesn't take place in a virtual underground shaft, but rather inside what is known as the "cloud": a centralized network made up of hundreds of thousands of servers, each storing staggering amounts of data. The term "cloud" generally refers to the technology, information, resources, and services provided on demand over the Internet, and it also alludes to systems that provide online computing instances and computing capacity from data centers. Cloud computing (the online development and utilization of computer technology) swelled into an estimated \$36 billion market in 2008, representing roughly 13% of global software sales.

Whether we are aware of it or not, all Internet users currently operate within the cloud. Every time we check our bank balances online, purchase an item in an online store, comment on a blog, upload a video to YouTube, or update our Facebook status, we are in the cloud. (Such services are what entrepreneur, author, and Web 2.0 guru Tim O'Reilly refers to as "cloud-based end-user applications.")

As you try to envision the cloud, imagine it being almost like a living

thing. It regenerates as it grows, expanding and swelling, collecting more and more information. People will increasingly enter this space through PCs, smart phones, PDAs, and other such devices. The mobile phone in your pocket, the laptop in your briefcase, the navigation system in your car—all of these are enveloped by, and connected to, the cloud. And instead of only being accessible at home or at work, your vital data can now be accessed on wireless portable devices as well.

The cloud represents a fundamental shift in how information is handled. At a basic level, the cloud's formation is analogous to the evolution of the electric grid, when farms and businesses began to purchase power from efficient industrial utilities rather than using their own generators. Similar to electricity, processing power and storage space can now be had cheaply and instantly. Also, most cloud computing installations are run on open-source software platforms. It used to be that data was kept separate, but now servers are merged. By operating in the cloud and creating digital-management systems capable of handling unlimited data, businesses and industries are cutting costs and operating expenses. It is like the early stages of community formation in that we cannot yet know for sure where power and authority will ultimately reside.

For example, say that a hedge-fund manager uses Google Spreadsheets to organize a list of employee Social Security numbers. The responsibility of maintaining and protecting that information from data breaches falls not into the fund's hands, but into Google's hands.

But clouds can also reduce expenses and expand employee access to computer systems. Departments within an organization would also be able to share a single database, thereby eliminating many of the barriers that separate siloed organizational functions, such as sales, marketing, HR, or billing. As these clouds expand and swell, an entirely new community of Internet start-ups will have access to enormous power. This will help level the playing field for small companies, and entrepre-

neurs will be able to start businesses with little to no investment capital. Similar to the dot-com bubble in the mid-1990s, Web 2.0-based companies will spring up around these new models—only this time (hopefully) they will be more efficient.

## YOUR OWN PRIVATE CLOUD

In the future, a host of customized devices, equipped with environmental sensors that send and receive data, will be able to tap into the power of the cloud. People will increasingly look to the cloud to engulf them in information on everything they need on a momentary basis, from restaurant menus to personal records.

New aggregation services are already attempting to consolidate the information from the cloud into customizable streams. One such service, called FriendFeed, allows users to funnel all of their online activities into a single broadcast and send it off to anyone who wants to be updated. FriendFeed also allows users to build a customized data feed composed of content that their online friends have shared across a broad range of social networks (such as posting on blogs and microblogs). Intel's more sophisticated Mash Maker allows people using their ordinary Web browsers to create "mash-ups"—Web application hybrids that combine content or data from different sites in unique and sometimes groundbreaking ways.

## CLOUD SECURITY

It is vital to look at the privacy-eroding aspects of these tools. Learning how to manage online privacy issues will be critical in the twenty-first century, especially as online personal data is controlled less and less by the person it belongs to. Already, people search engines can uncover social networking profiles, addresses and telephone numbers, online photographs, marriage and divorce records, criminal records, and much more. Even your movements can be tracked with cell phones equipped with Bluetooth. And Google envisions a future when it is able to advise people on what jobs to take and

## CLOUD MINING IN ACTION

According to market-research group IDC, approximately 4% of the companies they surveyed already use cloud computing in some form, and by 2012, that percentage will likely increase to 9%.

Trucking company Kenworth, for example, turned to cloud computing when looking for ways to cut down on its trucks' gas consumption. Engineers discovered that the mud flaps on its trucks were a major source of drag. Their research indicated that shortening and tapering the flaps would save about \$400 from a typical truck's annual gas bill.

To solve its mud flap dilemma, the company rented time on supercomputers, which were accessed via the Internet, to help it design its latest truck, the T660. The system helped employees identify various gas-guzzling design flaws they might have missed had they solely relied on computers at Kenworth's own facilities. The new aerodynamic designs have the potential to save truck fleets millions of dollars in gas costs each year. Kenworth used cloud computing to help streamline their operations and save money.

A leading online video provider, Ooyala, has also seen record growth, thanks in part to cloud computing. Ooyala has leveraged cloud services from Amazon Web Services, which has helped it pass along cost savings to its customers. Ooyala's cloud-based features offer companies and customers "pay as you go" and flexible payment options for video delivery and storage services while avoiding large upfront costs. Since online video requires significant storage, bandwidth, and computing power, cloud computing technology helps process thousands of hours' worth of video content much more quickly and efficiently.

—Erica Orange

how they might enjoy spending their days off.

The blending of Internet and mobile technologies is redefining and reconfiguring product and service offerings, especially as we rapidly gain more insight into individual interactions relative to location and time. Location-tracking technology, for instance, may be increasingly utilized by corporate sales teams, HR, and management as a way to learn more about both consumer and employee behavior. It would also enable organizations and third party providers to obtain even more data on a generation of youth that is disarmingly comfortable with posting a great deal of personal information online.

Reality mining may give managers unwarranted confidence that they can accurately and consistently predict what people will do. Organiza-

tions must be aware that they could face a destructive backlash from consumers and regulators if it is felt that location-tracking is being abused by employers, advertisers, marketers, and insurance companies (not to mention criminals, stalkers, law enforcement agencies, and the government).

Digital data, easily collected, can be stored indefinitely and reproduced and distributed with ease, almost without cost, inside of the cloud. People will have to accommodate themselves to the idea that their lives will be highly documented and that records provided both knowingly and unknowingly are part of a global digital future. Youthful indiscretions will follow individuals far into their lives as public records are digitized and made more accessible. Organizations and governments that deal with all this information will

have to increase efforts to avoid compromising the trust relationship they develop with their customers and constituents.

Concerns about terrorism may also result in more government pressure to virtually monitor employees, or even report on employee online activities. This massive cluster of networked data will eventually become so pervasive and powerful that it will respect no physical boundaries—and perhaps no current laws. Once this information is stored in one or more cloud-based databases, it may be accessed and used in ways that individuals never envisioned or intended. The problem is that information stored in the cloud, as opposed to in your office or data center, is not considered your private property. There is also little oversight. Governments can access this treasure trove of information with a subpoena, and companies can easily mine this information to deliver targeted advertising or marketing messages, and share with others.

It is also very difficult to remove data once it is uploaded. A 2008 report, written by the Constitution Project, a bipartisan coalition of First Amendment and civil liberties groups, suggests that executive and legislative actions are needed to lessen the privacy risks associated with cloud computing.

It has been said before that technology evolves along with, and in reaction to, social norms. But as Internet and mobile technology advances exponentially, perhaps the reverse is true. As we learn more about our own behavior, will social norms ever fully be able to catch up with what technology is capable of providing? Either way, it's clear that the political, social, and economic spheres are now being radically redefined by an ever-expanding cluster of networked data. □



### About the Author

Erica Orange is the senior analyst at Weiner, Edrich, Brown, Inc., a leading futurist consulting group in the United States. Her address is 200 East 33rd Street, Suite 9I, New York,

New York 10016. E-mail [erica.orange@gmail.com](mailto:erica.orange@gmail.com).