We live our lives in the present, recollect experiences from the past, and anticipate events in the future. At the same time, Einstein suggested that time is an illusion; that the past, present, and future form a four-dimensional block — that there’s no single special present and that all moments are equally real. Whichever theory better describes reality, we can safely say that heretofore little information about the future, not much of the present, and less of the past has been computationally accessible to us — but that’s changing.

Although we say life is precious, historically, most of us don’t keep detailed records of our past. I can’t tell you what I had for breakfast last Tuesday. Nor can I recall my child winning an important award that I find in a 25-year-old box in the garage. But we’re now at a great divide. While our ancestors and our younger selves left behind few and sparse records, we and the digital services we use are now capturing, preserving, and leveraging increasingly dense digital traces. Business analytics uses these traces to monitor, measure, model, and predict day-to-day activities, preferences, and behaviors, and personal analytics are taking off.

Key Challenges
This special issue of *IEEE Internet Computing* explores the “Internet of You.” In our call for papers, we requested articles that explored both Small Data (individual-scale data sources, processing, and modeling) and Big Data (community-level aggregation and analytics) on these potential topics:

- diverse data sources and digital traces, including email, Facebook, financial, health, location, images, sound, consumer transactions, and interests;
- methods to combine trace data into complete models; data liberation; diverse user models, such as the physical self, memories, aspect-limited versus comprehensive models; and data quality, including managing history, change, comprehensiveness, and accuracy;
- methods to aggregate and process heterogeneous datasets, stages-of-life ontologies, the scope and
purpose of these data collections, available public data sources;
• usage models for experience sampling — proximity, context, activity sensing, quantified self, situation-aware modeling, activities of daily living, my immortal avatar, workflows, and pattern learning;
• representation technologies such as agents, smartphones, wearable computing, personal sensing networks, pattern representation and adaptation, and natural language;
• new kinds of applications that draw insights from data analytics — including recommendation systems, personalized health, real-time marketing, and predicting elections from Twitter feeds;
• open architectures for personalization, the role of cloud computing, and relevant emerging standards;
• concerns regarding privacy and surveillance, the extent of privacy erosion, a taxonomy of privacy threats, and incentives and disincentives for sharing, the right to forget, and status of legal safeguards;
• privacy and security technology safeguards, including identity management, disclosure control, privacy-preserving data mining, de-identification, new security models, and mechanisms that audit and control personal information flows and usage; and
• social and philosophical implications for humans’ conception of self.

We received several quality submissions in response, and the articles we chose provide windows into these topics.

In This Issue
Available approaches for recording day-to-day experiences include diaries and blogs for recording thoughts on topics; email repositories for recording correspondence; the Wayback Machine for providing snapshots of websites over time; social networks such as Facebook where individuals record meaningful experiences to share with their 100 closest friends (for example, “walking to class”); browser histories; cell phones that capture location traces of their owners; life-logging wearable computing (such as Google Glass) where individuals automate data capture by wearing a camera or microphone that frequently takes pictures or records sounds of their daily lives; the quantified “self movement,” which aims to keep track of your heart beat, steps walked, and blood sugar level; and standardized electronic patient health records. In the first article of this issue, “My Digital Life: 2003 Onward,” Alberto Frigo describes a variant of life-logging, in which he augments automated data entry with annotations on thoughts and feelings. His is a highly personalized experience report, describing a system that the author has developed to document his daily life over the last 12 years via photographs, other sensor data, and commentary.

Of course, individuals aren’t alone. When we share our tweets or other datasets with social networks via the cloud, third parties can analyze this data to learn more about us, individually and collectively. Some have analyzed tweets to predict elections in real time; others have used accident reports and slowdowns to learn about speed limits and traffic snarls or help real-time responders assess emergency situations. As an example, in the second article, “Microbloggers as Sensors for Public Transport Breakdowns,” Mariluz Congosto and her colleagues describe the Metroaverias system for processing tweets to mine information about a specific problem: detecting breakdowns in public transportation in the Madrid metro system. Of particular interest are the classification, event detection, and visualization systems — along with the use of Twitter — to share information with the community. The system is also interesting because it uses semantic processing to geolocate tweets, even though only a small percentage had location information from devices. This article adds value in that it’s representative of work that ties individuals’ tweets back to aggregated and actionable community views.

In Star Trek, the Borg is an alien race that shares a hive mind so that, if an individual learns something, it can be shared with the collective (remember: “Resistance is futile!”). One aspect of creating a consistent shared model is trust, and it’s important to understand the properties and evolution of trust. In the third article, “Trust Networks: Topology, Dynamics, and Measurements,” Santa Agresti and her colleagues analyze two real-world social networks (Ciao and Epinions) with regard to trust relationships using graph analysis. They conduct empirical analysis of two trust networks from three perspectives: structural property, node importance ranks, and dynamics. The article shows that graph properties of the two social networks are similar and argues that some of the graph properties relate to trust relationships.
Finally, there's an increasing need in science and business to discover and manage heterogeneous distributed data sources (both big and small). The ontology problem of meshing overlapping datasets that differ in naming, detail, and other aspects makes semantic interoperability among datasets problematic. In the fourth article, "Building a Real-Time Web Observatory," Ramine Tinati and his colleagues describe a uniform, application-agnostic middleware architecture that the authors (and a community of others) call a Web observatory for discovering and managing Web-based data resources and stream data. Data authors can publish data sources; others can discover them; and the system provides querying and security middleware services. With a mix of manual and automated transformations, the Web observatory system can transform the source data's representation to follow the target user's metamodel.

The challenges related to information about humans are considerable, and these articles offer a glimpse into some of the problems—but because of the number of challenges, they barely scratch the surface. There are many kinds of life-logging with many varied purposes, so there's no overarching framework into which they all fit. More work is needed on trust—to better understand loss of trust, scope-creep where information gathered for one purpose is used for ever-widening purposes, and whether competition is possible in perfect sharing environments (for example, can airlines collude—effectively price sharing—by sharing their prices in public in real time?). The range of applications related to information analytics seems unlimited, especially as more personal information is shared in the cloud. And there’s a constant challenge to evolve middleware architectures to make it easier for more and more data discovery, sharing, and analysis.

At his trial, Socrates said that the unexamined life is not worth living. Others have quipped that an over-examined life can’t be enjoyed. How complete a model can we create? What about the Hawthorn effect of the collection act itself? One aspect of the scope is whether the records we collect stop with people. Surely, we also want to collect and analyze records on their pets, their vehicles, and pretty much any object in the world (for instance, truckers want to track many properties of truck tires). Do we archive all this forever (because we can probably afford to and it takes effort to semantically garbage-collect data)? If we have many such data collections, there are ontological puzzles of how to compose the data collections and how to derive new kinds of information from different collections. What standards will be needed? And of course, there's the question of who can see what data? Even at present, users are only vaguely aware of how much personal data are collected by third parties.

Then there are future applications. Technology already exists to track people using their mobile devices, record what they see and hear, recognize objects visually, autonomously navigate, recognize some kinds of gestures, and answer some kinds of spoken questions. Work on recognizing human workflows may provide a missing key to record what we’re doing, so computational systems can observe, make sense of, record, and answer questions about our heart operations or our activities of daily living. Today’s genealogy research focuses on gleaning facts about individuals from the past. But some communities are enthralled with the prospect of creating immortal avatars that provide a record of everything about an individual in a model that's queriable, conversant, and adaptive. Right now, there may be more challenges than answers. But there's tremendous momentum—some would argue manifest destiny—that will continue to drive research in this area.

Deborah Estrin is a professor of computer science at Cornell Tech in New York City, and she’s also a co-founder of the nonprofit startup, Open mHealth (www.openmhealth.org). Her current focus is on mobile health and Small Data, leveraging the pervasiveness of mobile devices and digital interactions for health and life management (see her TEDMED talk at www.youtube.com/watch?v=1AEhG5yEHWU). Estrin has a PhD in electrical engineering and computer science from the Massachusetts Institute of Technology. She’s a member of the American Academy of Arts and Sciences and the National Academy of Engineering. Contact her at destrin@cs.cornell.edu.

Craig W. Thompson is a professor emeritus in the Department of Computer Science and Computer Engineering at the University of Arkansas. His research interests include artificial intelligence, databases, middleware architectures, virtual worlds, RFID, and pervasive computing. Thompson has a PhD in computer science from the University of Texas at Austin. He’s an IEEE fellow. Contact him at cwt@uark.edu.