MAPPING CANCER'S GENES • HOW COLOR TRICKS THE BRAIN

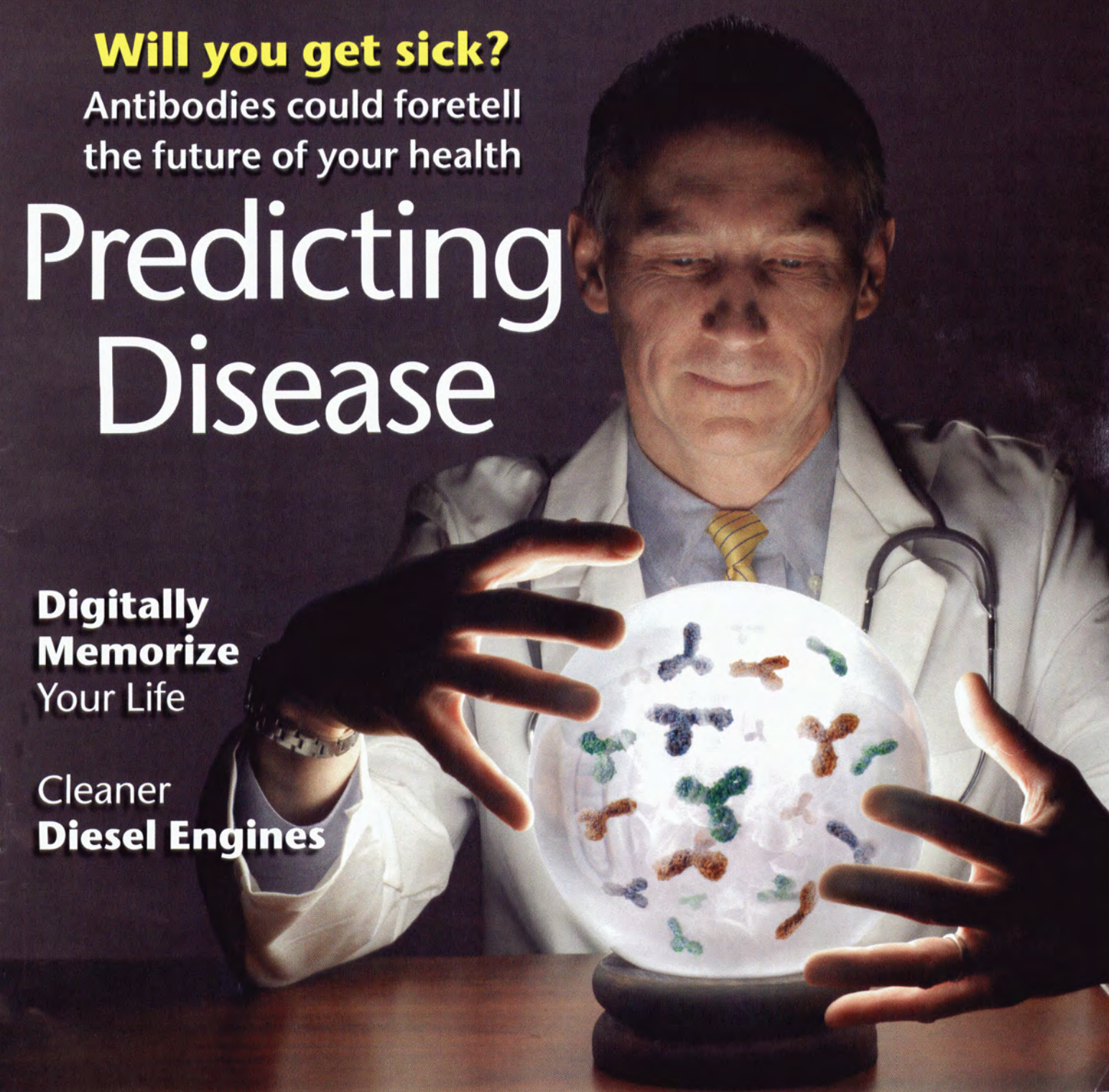
SCIENTIFIC AMERICAN

BLACK HOLE BLOWBACK:
Building Galactic Clusters

MARCH 2007 $4.99
WWW.SCIAM.COM

Will you get sick?
Antibodies could foretell the future of your health

Predicting Disease



Digitally Memorize Your Life

Cleaner **Diesel Engines**

"He often needed a portable, infallible, artificial memory."

New systems may allow people to record everything they see and hear—and even things they cannot sense—and to store all these data in a personal digital archive

A Digital Life

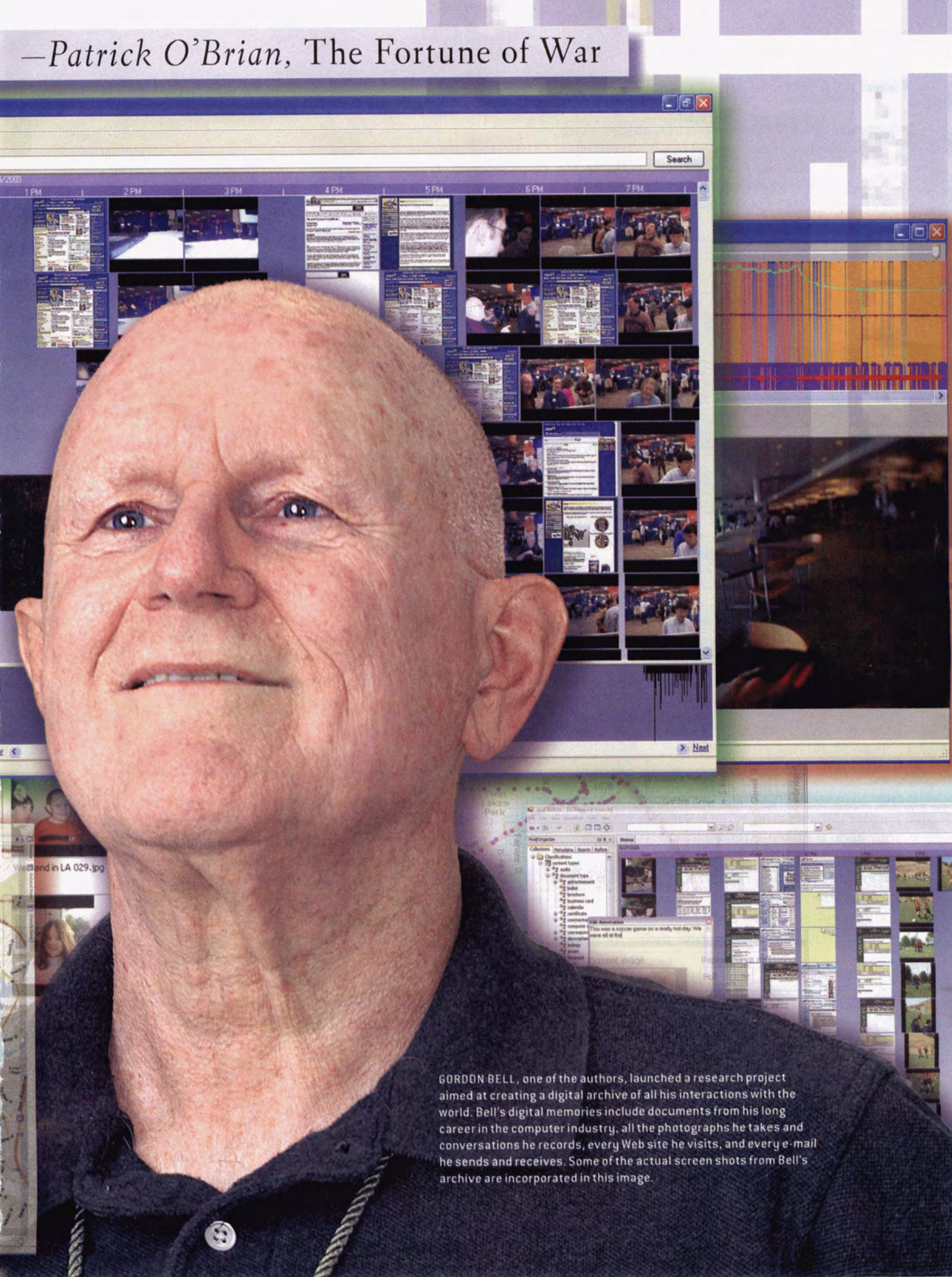
By Gordon Bell and Jim Gemmell

Human memory can be maddeningly elusive. We stumble upon its limitations every day, when we forget a friend's telephone number, the name of a business contact or the title of a favorite book. People have developed a variety of strategies for combating forgetfulness—messages scribbled on Post-it notes, for example, or electronic address books carried in handheld devices—but important information continues to slip through the cracks. Recently, however, our team at Microsoft Research has begun a quest to digitally chronicle every aspect of a person's life, starting with one of our own lives (Bell's). For the past six years, we have attempted to record all of Bell's communications with other people and machines, as well as the images he sees, the sounds he hears and the Web sites he visits—storing everything in a personal digital archive that is both searchable and secure.

Digital memories can do more than simply assist the recollection of past events, conversations and projects. Portable sensors can take readings of things that are not even perceived by humans, such as oxygen levels in the blood or the amount of carbon dioxide in the air. Computers can then scan these data to identify patterns: for instance, they might determine which environmental conditions worsen a child's asthma. Sensors can also log the three billion or so heartbeats in a person's lifetime, along with other physiological indicators, and warn of a possible heart attack. This information would allow doctors to spot irregularities early, providing warnings before an illness becomes serious. Your physician would have access to a detailed, ongoing health record, and you would no longer have to rack your brain to answer questions such as "When did you first feel this way?"

Our research project, called MyLifeBits, has provided some of the tools needed to compile a lifelong digital archive. We have found that digital memories allow one to vividly relive an event with sounds and images, enhancing personal reflection in much the same way that the Internet has aided scientific investigations. Every word one has ever read, whether in an e-mail, an electronic document or on a Web site, can be found again with just a few keystrokes. Computers can analyze digital memories to help with

—*Patrick O'Brian*, The Fortune of War



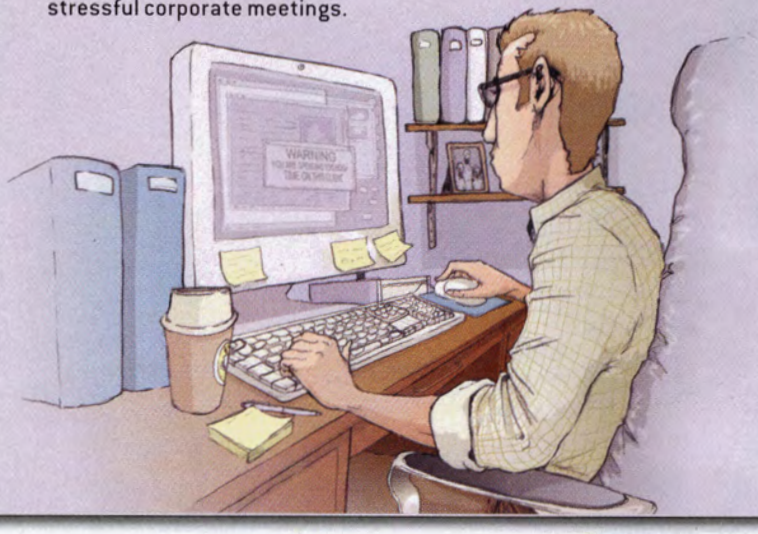
GORDON BELL, one of the authors, launched a research project aimed at creating a digital archive of all his interactions with the world. Bell's digital memories include documents from his long career in the computer industry, all the photographs he takes and conversations he records, every Web site he visits, and every e-mail he sends and receives. Some of the actual screen shots from Bell's archive are incorporated in this image.

A DAY IN THE DIGITAL LIFE

Access to lifelong digital memories can improve job productivity, medical treatment, school performance and much else. A glimpse of the daily routines of the Digitals, a fictitious family living in the not so distant future, illustrates some of the potential benefits.

Anne Digital, a chemistry professor, needs to review an article written by a colleague but cannot recall the name of the author. Anne does, however, remember viewing the document during a phone call with a student last summer. She limits her document search to those viewed during phone calls with the student—which were automatically logged at the time—and immediately sees the one she needs.

Dave Digital, a stockbroker, archives all his documents, e-mails, phone calls and Web site visits over the course of the workday. As Dave composes an e-mail, his time management program warns that he is spending too many hours communicating with an unimportant customer. He forgoes polishing the message and sends it off as is, then goes to work on a higher-priority account. Dave also reviews recent records of his weight, heart rate and caloric input to determine the negative effects of two days of stressful corporate meetings.



Dave takes his seven-year-old daughter, Laura, to the doctor. Laura's health analysis program recommended a checkup because her weight gain over the past six months (as logged daily by her bathroom scale) is lower than expected. The doctor says this symptom could be a side effect of Laura's asthma medication. Because her breathing has been excellent for several months, the doctor suggests taking her off the medication for a while.

time management, pointing out when you are not spending enough time on your highest priorities. Your locations can be logged at regular intervals, producing animated maps that trace your peregrinations. Perhaps most important, digital memories can enable all people to tell their life stories to their descendants in a compelling, detailed fashion that until now has been reserved solely for the rich and famous.

A Web of Trails

THE VISION of machine-extended memory was first expounded at the end of World War II by Vannevar Bush, then director of the U.S. government office that controlled wartime research. Bush proposed a device called the Memex (short for "memory extender")—a microfilm-based machine that would store all of an individual's books, records and communications. The Memex was to be built into a desk and equipped with a keyboard, a microphone and several display surfaces. The person behind the desk could use a camera to make microfilm copies of photographs and papers or create new documents by writing on a touch-sensitive screen. The Memex user could also mount a camera on his or her forehead to capture pictures while away from the desk. One of the most prescient of Bush's ideas was the suggestion that the Memex should be designed to imitate the associative thinking of the

Overview/*Digital Memories*

- Because human memory is fallible, researchers are striving to develop systems that can automatically record communications, documents, images and video, storing everything in a searchable archive.
- Ongoing advances in sensors and data storage promise to make digital recording easier. The bigger challenge is devising software that can organize the information.
- Digital memories may yield benefits in medical care, job productivity and other areas, but developers must ensure that the archives are secure.

MARK RICHARDS (*Gordon Bell portrait*); PHOTOCOMPOSITION BY JANA BRENNING (*preceding pages*); MATT VINCENT (*box above*)

At dinner Dave and Anne argue with their 14-year-old son, Steve. Anne is frustrated that Steve leaves his homework assignments to the last minute and wants him to start his next essay immediately. But Steve shows his parents the results of his education analysis program, which indicate that his grades are just as high when he does his homework late. Steve's digital memories also reveal that he is an auditory learner, benefiting more from group discussion than from reading.

Anne's mother, Jean, is a senior citizen living in an assisted-care home. The staff at the home have access to parts of Jean's digital memories; they will be automatically alerted if any heart or breathing irregularities occur or if her wearable monitors show that she is failing to take her usual walks. Anne has found that watching her mother's dishwasher log is insightful: when Jean feels run down, she often gets behind on the dishes. At bedtime Jean watches old photographs and videos from her digital archive, using an interactive display to take an extended walk down memory lane.

human mind, which he described in lively terms: "With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain."

Over the next half a century intrepid computer science pioneers, including Ted Nelson and Douglas Engelbart, developed some of Bush's ideas, and the inventors of the World Wide Web borrowed the concept of the "web of trails" to build their system of linking sites. But the Memex itself remained technologically out of reach. In recent years, however, rapid advances in storage, sensor and processor technologies have paved the way for new digital recording and retrieval systems that may ultimately go far beyond Bush's vision.

The growth of digital storage capacity has been staggering: today a $600 hard drive can hold a terabyte (one trillion bytes) of data, which is enough to store everything you read (including e-mails, Web pages, papers and books), all the music you purchase, eight hours of speech and 10 pictures a day for the next 60 years [*see table on page 64*]. If current trends continue, within a decade you will be able to carry the same amount of information in your cell phone's flash memory, while connecting wirelessly to a $100 four-terabyte drive on your PC. In 20 years $600 will buy 250 terabytes of storage—enough to hold tens of thousands of hours of video and tens of millions of photographs. This capacity should be able to satisfy anyone's recording needs for more than 100 years.

At the same time, manufacturers are producing a new generation of inexpensive sensors that may soon become ubiquitous. Some of these devices can record a wealth of information about the user's health and physical movements. Others can gauge the temperature, humidity, air pressure and light level in the surrounding environment and even detect the presence of warm bodies nearby. Some monitors are meant to be worn, and others are designed to be placed in rooms or incorporated into appliances such as refrigerators. (A fridge sensor could keep track of your snacking habits by measuring the number of times the door is opened.) And microphones and cameras are now cheap enough to be installed virtually anywhere—particularly in cell phones, where camera inclusion is becoming the norm,

DAILY JOURNEYS can be documented with the SenseCam (*hanging from Bell's neck in photograph above*), a camera that automatically takes pictures when its sensors indicate changes in light levels or the presence of people nearby. A Global Positioning System device continually tracks Bell's location, enabling the creation of visual diaries of his travels.

with voice recording soon to follow.

Finally, the dramatic increase in computing power over the past decade has led to the introduction of processors that can efficiently retrieve, analyze and visualize vast amounts of information. An ordinary notebook PC can run a database that is more powerful and almost 100 times as large as that of a major bank of the 1980s. An inexpensive cell phone can surf the Web, play videos and even understand some speech.

As the hardware for digital recording has improved, more and more people have started to create electronic chronicles of their lives. The advent of cheap, high-quality digital cameras (including those incorporated into cell phones) has triggered a boom in photography. Blogs that incorporate photographs are now becoming more common than personal Web sites. Young people in particular are embracing blogging and the use of mobile devices. The fact that this proliferation of digital chronicling is taking place with only very rudimentary tools demonstrates how deep the desire must be. And the interest will surely grow once the process of digital recording becomes easier and more comprehensive.

One Man's Memories

OUR OWN EXPERIENCE with digital memories began in 1998, when Bell decided to go paperless, doing away with a messy mountain of articles, books, cards, letters, memos, posters and photographs. To transfer this heap of memories to a digital record, Bell became obsessed with scanning all the documents and artifacts from his personal life and his long career in the computer business. (He even went so far as to scan the logos on coffee mugs and T-shirts.) He also began digitizing home movies, videotaped lectures and voice recordings. Bell is now paperless, but the cost was high: it took a personal assistant working for several years to complete the task. (Archiving more recent items has not required such strenuous effort, because the great majority of documents, images and videos are now created in digital formats, so capture is automatic.)

After scanning all this information, however, Bell became frustrated with his inability to make real use of it with the software available to him at the time. This frustration led to the MyLifeBits project. When we started the project in 2001, the search tools that had been developed for desktop computers were cumbersome. We set out to create a database that would give us the ability not only to do full-text searches of our PCs (a capability that is now commonplace) but also to quickly retrieve digital memories using attributes called metadata: for example, the date, place and subject of a photograph or written or spoken comments that the database appends to the file. Metadata are frequent-

MARK RICHARDS

THE AUTHORS

GORDON BELL and *JIM GEMMELL* have been working together on the MyLifeBits project at Microsoft Research since 2001. One of the pioneers of the computer industry, Bell oversaw the development of the famous VAX minicomputer for Digital Equipment Corporation in the 1970s. During the 1980s he became involved in public policy for computer science, and in 1995 he joined Microsoft as a principal researcher with its eSciences Research Group in San Francisco. Gemmell is a senior researcher in Microsoft's Next Media research group. His current focus is on personal lifetime storage, but his interests also include personal media management, telepresence and reliable multicast.

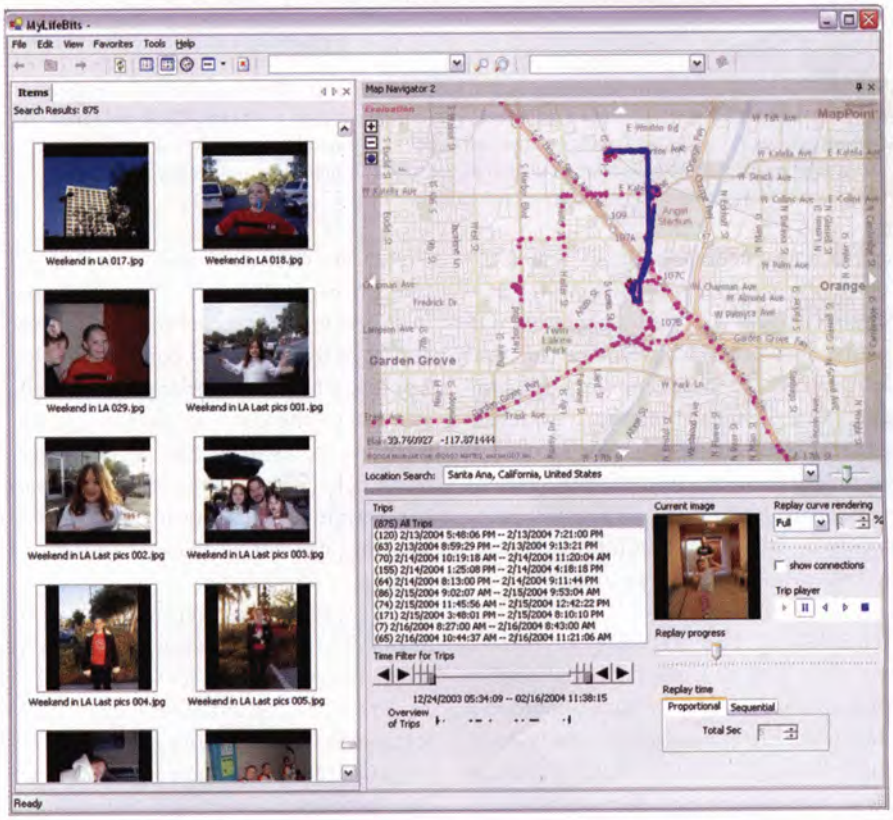
ly a crucial part of recall; a person seeking a specific e-mail, for instance, may remember that it was sent at a certain time of year. By linking these metadata, much of which are obtained automatically, to digital memories, the database allows users to efficiently comb through even the largest archives.

MyLifeBits has also provided Bell with a new suite of tools for capturing his interactions with other people and machines. The system records his telephone calls and the programs playing on radio and television. When he is working at his PC, MyLifeBits automatically stores a copy of every Web page he visits and a transcript of every instant message he sends or receives. It also records the files he opens, the songs he plays and the searches he performs. The system even monitors which windows are in the foreground of his screen at any time and how much mouse and keyboard activity is going on. When Bell is on the go, MyLifeBits continually uploads his location from a portable Global Positioning System device, wirelessly transmitting the information to his archive. This geographic tracking allows the software to automatically assign locations to Bell's photographs, based on the time each is taken.

To obtain a visual record of his day, Bell wears the SenseCam, a camera developed by Microsoft Research that automatically takes pictures when its sensors indicate that the user might want a photograph. For example, if the SenseCam's passive infrared sensor detects a warm body nearby, it photographs the person. If the light level changes significantly—a sign that the user has probably moved in or out of a room and entered a new setting—the camera takes another snapshot. A recent study led by researchers at Addenbrooke's Hospital in Cambridge, England, showed that a memory-impaired patient who reviewed SenseCam images every night was able to retain memories for more than two months. (In contrast, a nightly review of a written diary resulted in almost no improvement in memory retention.) Neuropsychologist Martin Conway of the University of Leeds in England speculates that the SenseCam could become "the first truly powerful 21st-century memory stimulant."

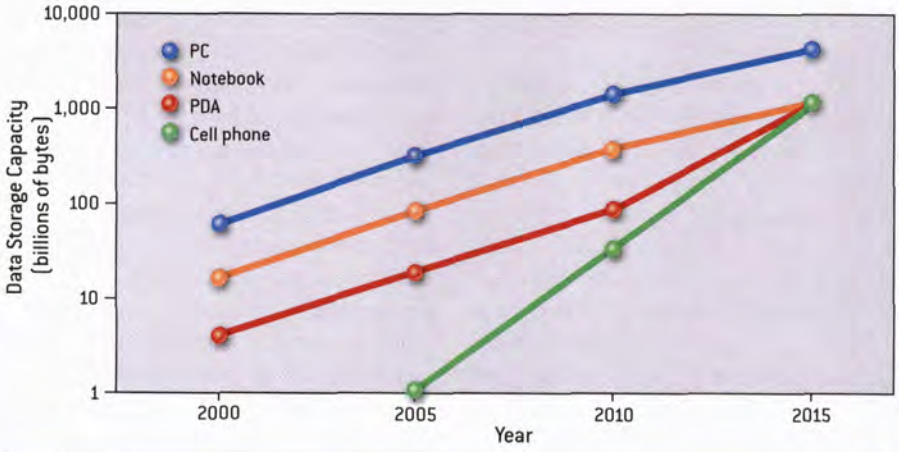
After six years, Bell has amassed a digital archive of more than 300,000 records, taking up about 150 gigabytes of memory. The information is stored on Bell's dual-disk notebook computer and his assistant's desktop PC, which are backed up locally and off-site. Video files grab the lion's share of the storage space—more than 60 gigabytes—whereas images take up 25 gigabytes and audio files (mostly music) occupy 18 gigabytes. The remainder is shared by 100,000 Web pages, 100,000 e-mails, 15,000 text files, 2,000 PowerPoint files, and so on. Bell has found the system particularly useful for contacting old acquaintances and finding other people with whom he needs to communicate. He has also employed MyLifeBits to retrieve Web sites for citations in his research papers, to provide doctors with records of a 25-year-old coronary bypass, and to obtain a photograph of a deceased friend for a newspaper obituary.

Some features of MyLifeBits, such as full-text search, have already been incorporated into commercial products. As a whole, though, the system requires more development to improve its ease of use and its management of the data. Better software for converting speech to text would greatly enhance the system by allowing users to search for words or phrases in phone conversations or other voice recordings. Similarly, automatic face recognition would solve the pesky problem of photograph labeling. And the retrieval of information could become easier if the system automatically identified the nature of each of the several hundred document types, perhaps by analyzing their form and content. But our research project has already dramatized the evolution of the PC from a word processor and number cruncher to a transaction processor that can log everything about the user's life in high-fidelity multimedia. Many experts have predicted the demise of the personal computer, but it is clear that the "P" in "PC" is not



FAMILY VACATIONS may become more memorable if they are digitally recorded and archived. Vacation snapshots are linked to GPS records (*pink dots on map*) in this diary of a trip to Los Angeles taken by Jim Gemmell, one of the authors, and his children.

COURTESY OF MICROSOFT RESEARCH



ITEM	SIZE EACH (BYTES)	DAILY RATE	ANNUAL STORAGE (BILLIONS OF BYTES)	60-YEAR STORAGE (BILLIONS OF BYTES)
Books/reports	1,000,000	1	0.4	21.9
E-mails	5,000	100	0.2	11.0
Scanned pages	100,000	5	0.2	11.0
Web pages	50,000	100	1.8	109.5
Purchased songs (MP3)	4,000,000	1	1.5	87.6
Recorded speech	1,000/ second	Eight hours	10.5	630.7
Photos (JPEG)	1,000,000	10	3.7	219.0
TOTAL			18.2	1,090.6

Numbers may not add up because of rounding.

STORAGE CAPACITY has grown tremendously in recent years and will continue to climb (*graph*). A typical desktop PC will hold more than a terabyte (a trillion bytes) of data by 2010; notebook computers, personal digital assistants and cell phones are expected to reach that milestone by 2015. A terabyte of storage is enough to hold all the books, e-mails, recorded conversations, songs and photographs that one is likely to accumulate over 60 years (*table*).

going away. If anything, PCs will become even more personal. What will change is the "C." Our machines will evolve into computing ecosystems that encompass not just computers but storage services on the Internet, new access devices (such as cell phones and entertainment centers), and ubiquitous sensors. Most likely our LifeBits will eventually be housed in a home server connected to various Web services.

Realizing the Vision

TO ILLUSTRATE the potential impact of digital memories, we have imagined a day in the life of a fictitious family making full use of this technology in the not so distant future [*see box on pages 60 and 61*]. Various pieces of the family's digital memories are stored in their personal devices—their cell phones, laptops, home computers and so on—but all that information is also securely transmitted over the Internet to a host server run by a hypothetical company called LifeBits, Inc. This company manages the storage of the data, performs regular backups (so as to recover any inadvertently deleted material) and places copies of the archive in various locations to ensure that it is not destroyed in a natural or man-made disaster.

Because most of their information is available via secure Web access, the family members can retrieve it anywhere and at any time. Particularly sensitive information that might put someone in legal jeopardy can be kept in an offshore data storage account—a "Swiss data bank," if you will—to place it beyond the reach of U.S. courts. The children in the family can encrypt their recordings, but the LifeBits service will give the parents access to the data in case of an emergency. Likewise, some of the parents' digital memories may be covered by employment contracts that stipulate that the data related to their jobs belong to their employers. When such employees leave their jobs, they may have to perform a "partial lobotomy" on their copies of the memories, expunging everything deemed to be company property.

Some of the scenarios we have described are not all that futuristic. Wearable sensor platforms that collect health data and monitor vital signs such as heart rate, breathing and the number of calories burned are already being commercialized by companies such as VivoMetrics in Ventura, Calif., and BodyMedia in Pittsburgh. In the meantime, Dust Networks in Hayward, Calif., has developed a wireless hub for relaying signals among a network of sensors. The Human Speechome Project, led by Deb Roy at the Massachusetts Institute of Technology Media Lab, is engaged in recording nearly every waking hour of the first three years of a child's life—the child is Roy's son, now a one-year-old—to study how people acquire language. And Kiyoharu Aizawa and his colleagues at the University of Tokyo are working on wearable video camera systems that would identify interesting moments to capture for posterity by monitoring the alpha waves in the user's brain.

Microsoft Research is supporting 14 universities undertaking a variety of projects in the field of digital memories. One of them is MyHealthBits, led by Bambang Parmanto of the University of Pittsburgh; this effort is taking on the challenge of recording huge amounts of health data and managing the voluminous records that result. Recent studies at the University of Washington have shown the benefits of continuous health monitoring in diabetic patients and individuals with sleep disorders.

This early progress is encouraging, but the advent of the digital-memories era will not be trouble-free. Some countries and U.S. states currently impose restrictions on recording conversations or photographing people. Many individuals are

JEN CHRISTIANSEN

WEARABLE MONITORS that constantly record and archive an individual's vital signs can enable doctors to detect illnesses before they become serious. For example, the SenseWear armband, produced by a Pittsburgh company called BodyMedia, can calculate the number of calories burned by measuring the skin's temperature, heat loss and impedance of electric current.

equally concerned about recording information that could be used against them in court. Digital memories, unlike those in our brains, would be fair game in a legal proceeding. Richard Nixon famously advised his aides to say "I can't recall" when testifying before a grand jury, but tape recordings of his own conversations were his downfall. For those of us who view digital memories as an extension of our own minds, the use of such materials in court would feel like self-incrimination. New technologies, however, can help minimize the potential dangers. When recording others, for instance, it may be possible to obscure their images or speech to avoid illegal recording.

Guarding the privacy of digital memories will be critical. The prospect that identity thieves, gossipmongers or authoritarian states could gain access to such records is frightening. Most people, however, already have quite a lot of sensitive information on their PCs. Security is an important concern regardless of how far you go with the concept of digital memories (although storing a lifetime of personal data in a single archive does at least make the problem quantitatively worse, if not qualitatively). Furthermore, even if our computer systems can be made as secure as Fort Knox, users must be very careful when sharing their information; with a single errant keystroke, one's medical records might inadvertently be distributed to the entire world. To prevent such mistakes, the user interfaces for digital memories must be better than the ones we have now, and we will need intelligent software to provide warnings when sharing data looks risky.

Another technical challenge will be ensuring that users are able to open their digital files decades after storing them. We have already run into cases where we could not access documents because their formats were obsolete. Digital archivists will have to constantly convert their files to the latest formats, and in some cases they may need to run emulators of older machines to retrieve the data. A small industry will probably emerge just to keep people from losing information because of format evolution.

An even bigger challenge will be devising software that can enable computers to perform useful tasks by tapping into this gigantic store of collected knowledge. The ultimate goal is a machine that can act like a personal assistant, anticipating its user's needs. At a minimum, computers must do a better job of organizing the information. Search strategies that work well for a few shelves of books may not work at all for a collection the size of the Library of Congress. Most of us do not want to be the librarians of our digital archives—we want the computer to be the librarian!

Consequently, our research group is very interested in applying artificial intelligence (AI) to digital memories. Although many experts are skeptical about AI efforts, we believe that such software may yield practical results if it can draw on the tremendous stores of data in personal archives. An AI system designed to work with a wealth of information is bound to perform better than one that has to make a recommendation based on very few data points. We have begun work in this area, developing software that could organize files based on their content, but much remains to be done.

In a sense, the era of digital memories is inevitable. Even those who recoil at our vision will have vastly more storage on their computers in the coming years and will expect software to help them more and more in utilizing it. Although some may be frightened at the prospect of ubiquitous recording, for us the excitement far outweighs the fear. Digital memories will yield benefits in a wide spectrum of areas, providing treasure troves of information about how people think and feel. By constantly monitoring the health of their patients, future doctors may develop better treatments for heart disease, cancer and other illnesses. Scientists will be able to get a glimpse into the thought processes of their predecessors, and future historians will be able to examine the past in unprecedented detail. The opportunities are restricted only by our ability to imagine them. SA

MORE TO EXPLORE

As We May Think. Vannevar Bush in *Atlantic Monthly*, Vol. 176, No. 1, pages 101–108; July 1945.

A Personal Digital Store. G. Bell in *Communications of the ACM*, Vol. 44, No. 1, pages 86–91; January 2001.

Digital Memories in an Era of Ubiquitous Computing and Abundant Storage. Mary Czerwinski, Douglas W. Gage, Jim Gemmell, Catherine C. Marshall, Manuel A. Pérez-Quiñones, Meredith M. Skeels and Tiziana Catarci in *Communications of the ACM*, Vol. 49, No. 1, pages 44–50; January 2006.

MyLifeBits: A Personal Database for Everything. Jim Gemmell, Gordon Bell and Roger Lueder in *Communications of the ACM*, Vol. 49, No. 1, pages 88–95; January 2006.

Information about MyLifeBits can be found at **www.mylifebits.com**