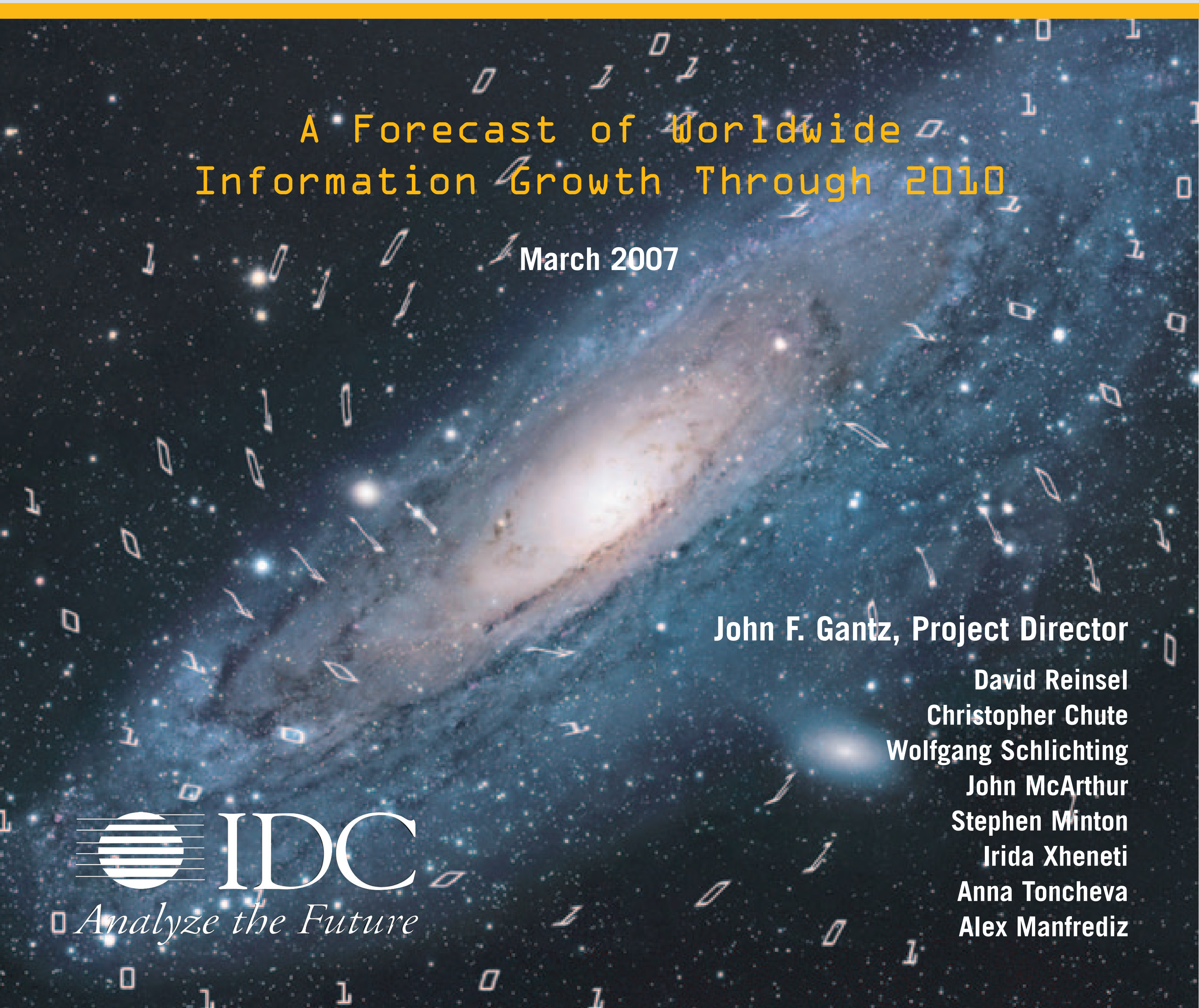
An IDC White Paper - sponsored by EMC

The Expanding Digital Universe



A Forecast of Worldwide Information Growth Through 2010

March 2007

John F. Gantz, Project Director

David Reinsel
Christopher Chute
Wolfgang Schlichting
John McArthur
Stephen Minton
Irida Xheneti
Anna Toncheva
Alex Manfrediz

IDC
Analyze the Future

EXECUTIVE SUMMARY

The airwaves, telephone circuits, and computer cables are buzzing. Digital information surrounds us. We see digital bits on our new HDTVs, listen to them over the Internet, and create new ones ourselves every time we take a picture with our digital cameras. Then we email them to friends and family and create more digital bits.

There's no secret here. YouTube, a company that didn't exist just a few years ago, hosts 100 million video streams a day.[i] Experts say more than a billion songs a day are shared over the Internet in MP3 format.[ii] Digital bits. London's 200 traffic surveillance cameras send 64 trillion bits a day to the command data center.[iii] Chevron's CIO says his company accumulates data at the rate of 2 terabytes – 17,592,000,000,000 bits – a day.[iv] TV broadcasting is going all-digital by the end of the decade in most countries. More digital bits.

What is a secret – one staring us in the face – is how much all these bits add up to, how fast they are multiplying, and what their proliferation imply.

This White Paper, sponsored by EMC, is IDC's forecast of the digital universe – all the 1s and 0s created, captured, and replicated – and the implications for those who take the photos, share the music, and generate the digital bits and those who organize, secure, and manage the access to and storage of the information.

Some of the key findings:

- In 2006, the amount of digital information created, captured, and replicated was 1,288 x 10^{18} bits. In computer parlance, that's 161 exabytes or 161 billion gigabytes (see sidebar). This is about 3 million times the information in all the books ever written.
- Between 2006 and 2010, the information added annually to the digital universe will increase more than six fold from 161 exabytes to 988 exabytes.
- Three major analog to digital conversions are powering this growth – film to digital image capture, analog to digital voice, and analog to digital TV.
- Images, captured by more than 1 billion devices in the world, from digital cameras and camera phones to medical scanners and security cameras, comprise the largest component of the digital universe. They are replicated over the Internet, on private organizational networks, by PCs and servers, in data centers, in digital TV broadcasts, and on digital projection movie screens.
- IDC predicts that by 2010, while nearly 70% of the digital universe will be created by individuals, organizations (businesses of all sizes, agencies, governments, associations, etc.) will be responsible for the security, privacy, reliability, and compliance of at least 85% of that same digital universe.
- This rapidly expanding responsibility will put pressure on existing computing operations and drive organizations to develop more information-centric computing architectures.
- IT managers will see the span of their domains considerably enlarged – as VoIP phones come onto corporate networks, building automation and security migrates to IP networks, surveillance goes digital, and RFID and sensor networks proliferate.
- Information security and privacy protection will become a boardroom concern as organizations and their customers become increasingly tied together in real-time. This will require the implementation of new security technologies in addition to new training, policies, and procedures.
- IDC estimates that today, 20% of the digital universe is subject to compliance rules and standards, and about 30% is potentially subject to security applications.
- The community with access to corporate data will become more diffuse – as workers become more mobile, companies implement customer self service, and globalization diversifies customer and partner relationships and elongates supply chains.
- The growth of the digital universe is uneven. Emerging economies – Asia Pacific without Japan and the rest of the world outside North America and Western Europe – now

account for 10% of the digital universe, but will grow 30%-40% faster than mature economies.

- In 2007 the amount of information created will surpass, for the first time, the storage capacity available.

This incredible growth of the digital universe means more than simply the fact that as individuals we will be facing information explosion on an unprecedented scale. It has implications for organizations concerning privacy, security, intellectual property protection, content management, technology adoption, information management, and data center architecture.

The growth and heterogeneous character of the bits in the digital universe mean that organizations worldwide, large and small, whose IT infrastructures transport, store, secure, and replicate these bits, have little choice but to employ ever more sophisticated techniques for information management, security, search, and storage.

HOW DID WE GET THE NUMBERS?

Information about our methodology and underlying assumptions can be found in the section "Methodology and Key Assumptions," but our basic approach was to take IDC forecasts for devices that create or capture digital information – personal computers, digital cameras, servers, sensors, etc. – and estimate the total number of megabytes they capture or produce in a year. We used IDC research and other sources to estimate how much of that data was replicated or copied – as email attachments, archived files, broadcasts, and so on.

Our research follows on previous work conducted at the University of California, Berkeley. Although our methodology varied from that in the Berkeley study – which examined the creation of original information (not including copies) and estimated how much digital information that would represent if all of it were converted to digital format – many of the underlying assumptions were the same.[v]

But our methodology allowed us to size and forecast all the information created and replicated in the digital universe, segment it by region, and put it in context with the available storage capacity. We believe ours is the first-ever study to size and forecast the rate of expansion of the entire digital universe.

WHAT ARE BITS AND BYTES?

A "bit" is the smallest unit of information that can be stored in a computer, and consists of either a 1 or 0 (or on/off state). All computer calculations are in bits.

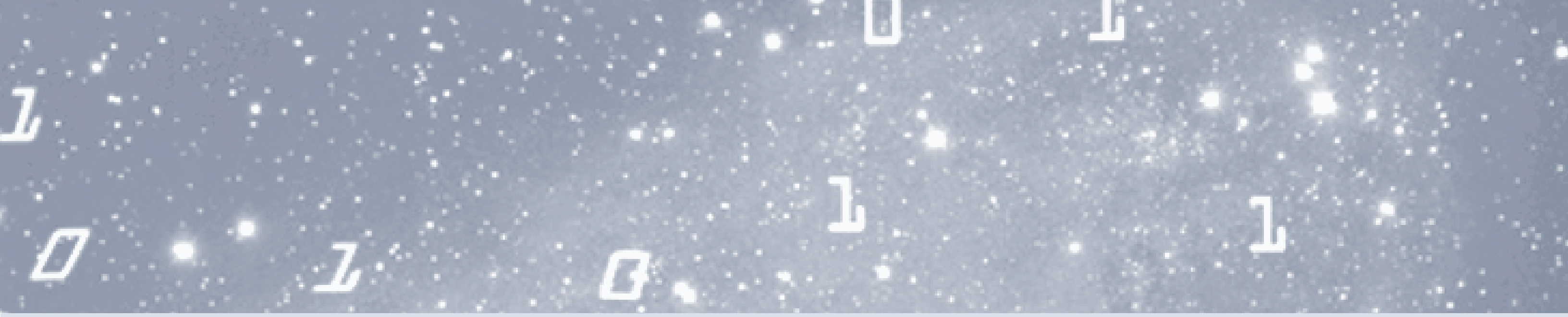
A "byte" is a collection of 8 bits. Bytes are convenient because, when converted to computer code, they can represent 256 characters, such as numbers or letters. So a byte is 8 times larger than a bit.

Common aggregations for bytes come in multiples of 1,000, such as kilobyte, megabyte, gigabyte, and so on. The progression is as follows:

Bit (b)	1 or 0
Byte (B)	8 bits
Kilobyte (KB)	1,000 bytes
Megabyte (MB)	1,000 KB
Gigabyte (GB)	1,000 MB
Terabyte (TB)	1,000, GB
Petabyte (PB)	1,000 TB
Exabyte (EB)	1,000 PB
Zettabyte (ZB)	1,000 EB

This seems simple enough, except sometimes multiples of bytes are considered as powers of 2, since the original machine language only has two states, 1 or 0. A kilobyte would then be 2^{10} bytes, or 1,024 bytes. A megabyte would be 2^{20} bytes, or 1,024 kilobytes, and so on.

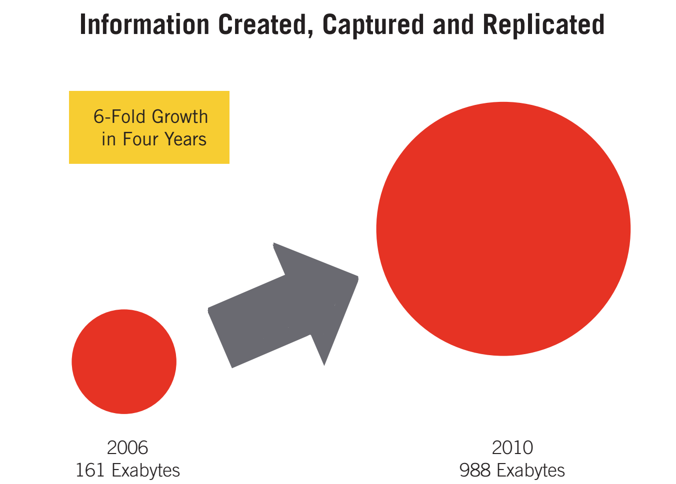
For the sake of simplicity, in all calculations for this research we used the decimal system we mentioned first. This is consistent with the representation used in the Berkeley study.



HOW BIG IS THE DIGITAL UNIVERSE?

The IDC sizing of the digital universe – information that is either created or captured in digital form and then replicated in 2006 – is 161 exabytes, growing to 988 exabytes in 2010, representing a compound annual growth rate (CAGR) of 57% (Figure 1).

Figure 1



Source: IDC, 2007

About one quarter of the digital universe is original (pictures recorded, keystrokes in an email, phone calls), while three quarters is replicated (emails forwarded, backed up transaction records, Hollywood movies on DVD).

A majority of these bits represent images, both moving and still. This is because one digital camera image can generate a megabyte or more of digital information, and video or digital TV can generate a dozen megabytes per second.

Voice signals, on the other hand, can be carried at less than one megabyte a second; and it would take a good typist more than a day and a half to produce a megabyte of keystrokes.

Although many of the images created are by individuals, they enter an organization's domain in email systems, in Web postings, and in applications from medical imaging and public safety surveillance to compound documents supporting insurance claims, recorded Web conferences, and advertising and marketing content.

To give you an idea of where all these exabytes come from, just consider the number of devices or subscribers in the world that can create or capture information.

Here is a partial list:

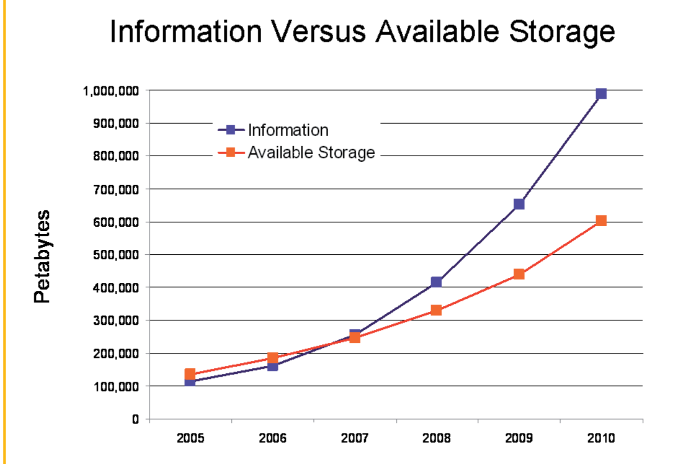
Category	Millions in 2006
Digital Cameras	400
Camera Phones	600
PCs	900
Audio Players	550
Mobile Subscribers	1,600
LCD/Plasma TVs	70

By 2010 this installed base of devices and subscribers will be 50% larger, devices will be cheaper, and resolutions higher. All creating more and more digital bits.

How much of the information that is captured, created, or replicated also is stored is another matter. As part of the research for this project, IDC also looked at how much storage will be available to store all this information, should we choose to.

Figure 2 shows the relationship of information created and storage capacity available on various storage technologies.

Figure 2



Source: IDC, 2007

HOW BIG IS THE DIGITAL UNIVERSE, REALLY?

It is pretty easy to picture a byte – it's the equivalent of a character on a page – or even a megabyte, which contains about the same amount of information as a small novel. But what about a million million megabytes, which is an exabyte?

If we stick with the book analogy, then the digital universe in 2006 could be likened to 12 stacks of books extending from the Earth to the sun. Or one stack of books twice around the Earth's orbit. By 2010 the stack of books could reach from the sun to Pluto and back. In 2006 those books would represent about 6 tons of books for every man, woman, and child on Earth. A large adult elephant weighs about 6 tons.

Still hazy on how big the digital universe is? In 2006 if you printed out all the exabytes onto typewritten pages, you'd have enough paper to wrap Earth four times over.

However, at the same time the digital universe is growing rapidly, bits and bytes themselves are getting smaller. That is, the circuits or media that store them are increasingly able to pack more into the same amount of space. In 1956, when IBM introduced the first disk drive, it could only store 2,000 bits per square inch, a measure commonly referred to as areal density. Today disks routinely store 100,000,000,000 bits per square inch. In the past, areal density growth of disks has been as aggressive as 100% per year. Over the last few years, and for the foreseeable future, areal density is expected to double every 2 – 3 years.

So, in a way, as the digital universe gets bigger, it is also getting smaller. That makes it even harder to visualize.

IDC is predicting that in 2007 the amount of information created and replicated (255 exabytes) will surpass, for the first time, the storage capacity available (246 exabytes). The storage media available to store the bits and bytes of the digital universe will grow 35% a year from 2006 to 2010, while the information created and replicated will grow by 57% a year in the same time period.

Not all of the bits in the digital universe will necessarily need to be stored - such as digital TV signals we watch but don't record, Web pages that disappear when we turn off our browser, or voice calls that are made digital in the network backbone for the duration of a call. On the other hand, we may want to store them. Personal video recorders and set-top boxes may store them temporarily, anyway; whether we program them to do so or not. And more and more of an organization's VoIP calls or Web site history may be recorded for legal reasons.

But whether this information gets stored permanently or not, it will be transported over networks, shuttled from switch to switch, stored temporarily somewhere, and otherwise require use of networking and storage infrastructures, both those in organizations and those in carriers, hosting firms, and other digital information service providers.

THE GROWTH OF ORGANIZATIONAL INFORMATION

Growing even faster than the digital universe as a whole is the subset created and replicated by organizations. In 2006, about 25% of the bits in the digital universe were created or replicated in the workplace; by 2010 that proportion will rise closer to 30%. (The rest of the universe will be mostly music, videos, digital TV signals, and pictures.)

Factors driving the growth of information in organizations include the increased computerization of small businesses, regulations mandating new archiving and privacy standards, and industry-specific applications – from security imaging and Internet commerce to medical imaging, sensor networks, and customer support applications that now include Web-based "click-to-talk" service.

Consider Wal-Mart, reputed to have the largest database of customer transactions in the world. In 2000, that database was reported to be 110 terabytes, with recordings and storage of information on tens of millions of transactions a day.[vi] By 2004, it was reported to be half a petabyte.[vii] Wal-Mart's data not only support internal decisions, but provide information to thousands of suppliers, as well.

(Imagine how many times each cash register keystroke is recorded and disseminated. By now, that Wal-Mart data and the bits replicated to other organizations could represent close to one percent of the digital universe.)

Or think of what oil companies call the "digital oilfield," a concept that calls for the integration of real-time production and drilling systems with reservoir modeling and simulation and that, as a by-product, generates a ton of data. A typical oil company might have 350 terabytes of data generated by 50 3D seismic projects, 10 terabytes in simulation models, 10 gigabytes a day coming in from oil field telemetry, and 4 terabytes a day of data tied up in 30,000 subnetworks at the refinery.

Although our research wasn't specific enough to segment organizational information by size of business or specific industry, we were able to estimate that three quarters of organizational information lies in the domain of the data center, another one quarter out in other departments. As we will see later, though, the responsibility for security, privacy protection, and compliance with legal requirements regarding data retention, is almost 100% centralized.

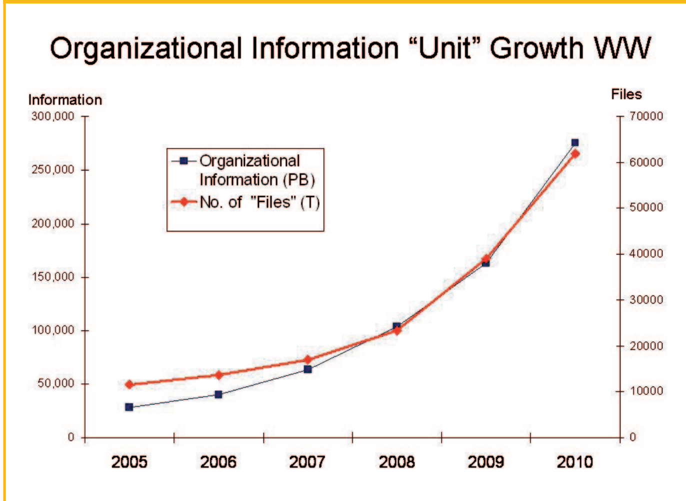
NOT JUST MORE INFORMATION, BUT MORE FILES

Over time, just as the total amount of information in the digital universe expands, so does the number of containers (e.g., electronic files, packets, digital images) for that information (Figure 3).

Even while image files grow to multi-megabyte size as a result of better camera resolution, the exponential growth of sensors, RFID tags, and packets created by IP voice phone calls is streaming trillions of smaller signals, some just 128 bits, into the digital universe.

While this may not seem like a big deal to some, it does impose an added burden on those who manage the bit streams of the digital universe, from Internet service providers and managers of backbone switches, to the IT managers who must deal not only with the management of larger quantities of information, but also more units of information, and more diverse types of information.

Figure 3



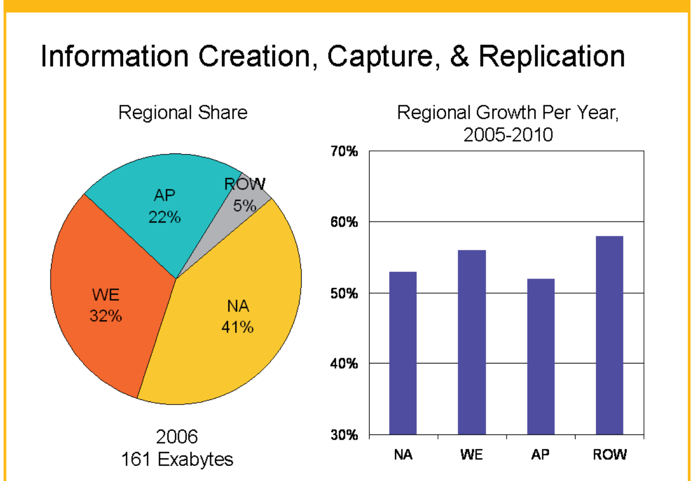
Source: IDC, 2007

THE REGIONAL PICTURE

The distribution of the expanding digital universe by geographic region more or less resembles IT spending by region. All regions are growing, although the emerging economies across the world and in particular in the Asia Pacific region are growing faster than the worldwide average (Figure 4).

This stable, if rapid, growth masks some underlying digital universe dynamics. In the mature economies of North America, Japan, and Western Europe, digital information growth is driven as much by increased device usage and resolution as by

Figure 4



Source: IDC, 2007

device penetration of the population as a whole.

In emerging economies, this dynamic is reversed. The growth of the digital universe is driven more by penetration of the devices into the population than by an increase in device capacities or resolutions.

The relationships between population penetration and IT intensity can be seen in the percentages in the table "Digital Universe Penetration Metrics" (Figure 5).

The Rest of World (ROW) sector and India and China together account for just 13% of IT spending but 69% of world population. The figure for Internet user share – 38% – sits between the two.

While we didn't segment the share of the digital universe by country, you would expect the share of the emerging economies to migrate from something close to IT spending to something closer to Internet usage.

We would estimate that the share of the digital universe attributable to emerging economies, including India, China, Eastern Europe, Latin America, the Middle East, and Africa sits today at close to 10% of the digital universe. That proportion will grow 30%-40% faster than the share of mature economies.

Some of the gating factors for these emerging economies will be how fast they convert their TV infrastructure to digital transmission, how many consumers can afford high end electronics, the rollout of sophisticated data-rich organizational applications, the automation of small business, and the deployment of surveillance cameras.

WHAT'S DRIVING GROWTH?

There are a number of trends at work creating this rapid expansion of the digital universe. These range from the growth of the Internet and broadband availability, to the conversion of formerly analog information – film, voice calls, TV signals – to digital format.

Falling prices and increased performance for digital devices, from phones and cameras to RFID tags and computers, also help drive up usage. So does the ability to store the information and share it in standard formats, such as MPEG 2, MP3, or MPEG 4.

The falling price of storage and processing power has also made industry adopt data-intense applications. The electronic "paperwork" behind the average insurance claim may now include several megabytes of digital pictures. Law enforcement and public safety organizations are rapidly adding digital security signals to their incoming data feeds, while police departments are experimenting with digital systems that scan license plates from cameras in police cars.

Meanwhile, the digital content of the average movie keeps increasing, and movie theaters themselves are starting to go digital. Graphics-intensive applications, from molecular modeling in pharmaceutical designs to visualization in automobile design and simulation are growing organizational databases. IDC research in 2006 indicated that almost one-fifth of organizations expect their data warehouses to double in 2007.

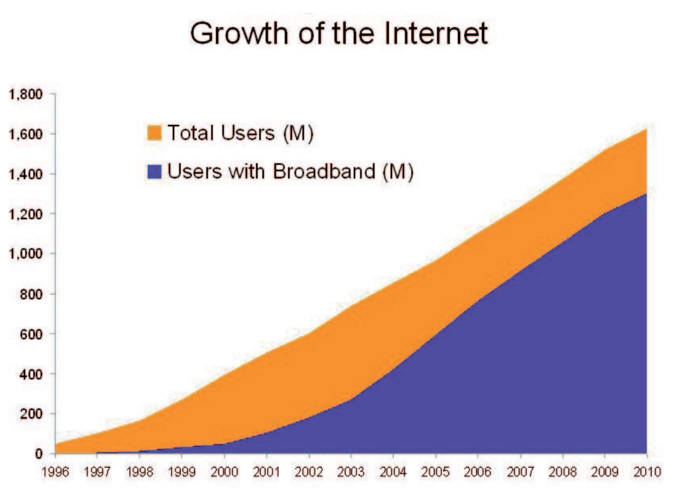
Figure 5: Digital Universe Penetration Metrics

	2006			2010		
	IT Spending	Internet Users	Population	IT Spending	Internet Users	Population
North America	41%	22%	5%	39%	16%	5%
Western Europe	31%	25%	7%	30%	22%	6%
Asia Pacific "Developed"	15%	15%	19%	13%	17%	19%
Rest of World (incl China, India)	13%	38%	69%	18%	45%	70%
	100%	100%	100%	100%	100%	100%

Source: IDC, 2007

But the prime mover may be the Internet. In 1996 there were only 48 million people routinely using the Internet. The Worldwide Web was just four years old. By 2006, there were 1.1 billion users on the Internet. By 2010 we expect another 500 million users to come online (Figure 6).

Figure 6



Source: IDC, 2007

At the same time the number of users with broadband access has also grown – and is expected to grow even more. Today over 60% of Internet users have access to broadband circuits, either at home or at work or school.

The rapid growth of the Internet – and more and more high speed access – has increased the ability of people to share and communicate information and their interest in doing so.

Take email. Since 1998 the number of email mailboxes has grown from 253 million to nearly 1.6 billion in 2006. Before the decade ends, the number of mailboxes is expected to taper off near 2 billion.

During the same period, 1998 to 2006, the number of emails sent grew three times faster than the number of people emailing – in part because of the growth of spam, and in part because people simply sent more emails. And surely the average corporate manager of email systems will tell you that messages are going out with more attachments and being stored longer (Figure 7).

Figure 7

Source: IDC, 2007

IDC estimates that in 2006, just the email traffic from one person to another – i.e., excluding spam – accounted for 6 exabytes (or 3%) of the digital universe.

THE IMAGE EXPLOSION

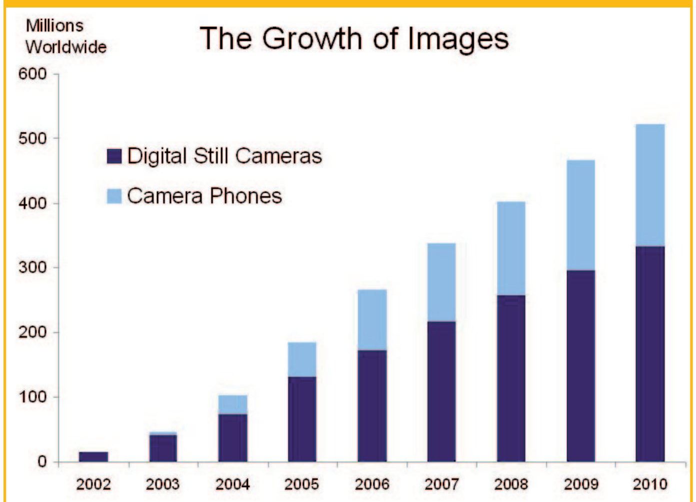
Between 2006 and 2010, the information added annually to the digital universe will increase more than six fold from 161 exabytes to 988 exabytes. One quarter of those exabytes will be images from cameras and camcorders.

The number of images captured on consumer digital still cameras in 2006 exceeded 150 billion worldwide, while the number of images captured on cell phones hit almost 100 billion. By 2010, IDC is forecasting the capture of more than 500 billion images (Figure 8). Each year the resolution of the pictures gets better and the megabytes per image grow.

Then there is video. In surveys conducted in 2006, IDC found that 77% of digital camera users had a video feature with their camera, and 50% of camera phone users had one. The feature was generally used 3-5 times a month, and each video clip tended to last from 30 seconds (camera phones) to a minute and a half (digital cameras).

But the real growth will come in camcorder usage – which should double in total minutes of use between now and 2010 – and digital surveillance cameras, which are expected to grow

Figure 8



Source: IDC, 2007

more than tenfold between 2006 and 2010 as analog systems are replaced by digital ones and as the number of total cameras installed increases.

SPEAKING OF VOICE

Another big sector of the digital universe – close to 20% of information created in 2006 – is voice. But because the number of minutes per call is not expected to grow appreciably between now and 2010 and because compression will get better, its share of the digital universe will drop considerably.

The big question mark for voice is replication. With no information on how many calls are actually recorded we simply opted for zero replication of original calls, but we did account for storage of information about the calls. We also estimated replication related to voice mail and storage associated with Voice over IP calls. Change this original assumption and the digital universe is even bigger than depicted.

THE USER AS PUBLISHER; THE ORGANIZATION AS CUSTODIAN

The Internet has created another aspect of the digital universe – the source of the majority of these bits. IDC estimates that of the 161 exabytes of information created or replicated in 2006, about 75% were created by consumers – taking pictures, talking on the phone, working at home computers, uploading songs, and so on.

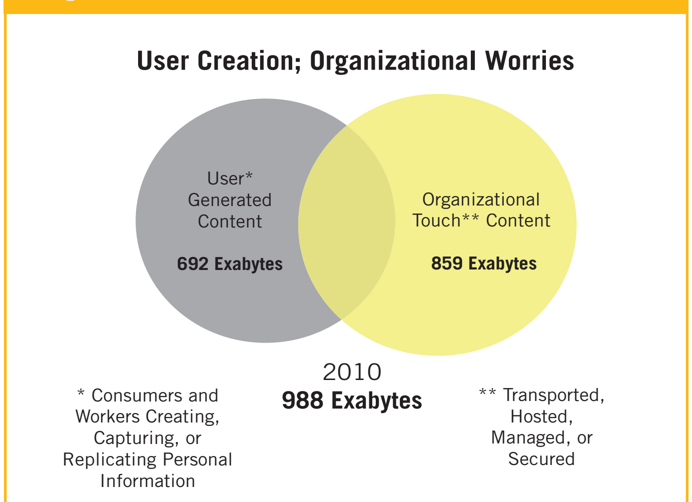
So enterprises only have to worry about 25% of the digital universe, right?

Not at all. Most user-generated content will be touched by an organization along the way – on a network, in the data center, at a hosting site, in a PBX, at an Internet switch, or a back-up system.

Consider camera phones, used by individuals at both work and play. Won't corporations have to worry about what pictures are being taken, messages sent, or purchases being made from these phones when they are used at work? Who owns contact lists? How will work-related phone emails be archived?

The left circle in Figure 9 shows a rough approximation of how much of the digital universe in 2010 will be created by individuals, meaning consumers and workers creating, capturing, or replicating information in the organization. In the right circle the figure shows how much of the digital universe will be touched – meaning managed, hosted, transported, or secured – by an organization.

Figure 9



Source: IDC, 2007

Corporate responsibility for information security and privacy will be tied not only to the information created by users in the digital universe, but also to information about that information.

In the digital universe, customer names and addresses, transaction records, account numbers, or search queries take up the merest fraction of total exabytes. They do, however, create a huge responsibility among enterprises to safeguard privacy and security, the breach of which can be a CIO's nightmare.

IDC believes that by 2010 while enterprises will create, capture, or replicate about 30% of the digital universe, they will have to worry about security, privacy, reliability, and compliance for more than 85% of it.

WHERE WILL WE STORE ALL THIS INFORMATION?

IDC forecasts that the media available to store the newly created and replicated bits and bytes of the digital universe will grow 35% a year from 2006 to 2010, or from 185 exabytes to 601 exabytes.

This forecast was created by adding newly shipped storage for any single year to an estimate of the storage still available on media shipped in previous years.

Figure 10 shows that growth by storage technology. The graph represents the storage capacity of each technology that is available to save new digital content in any given year. It does not represent where digital content resides – that is a different

STORAGE TECHNOLOGY: WHAT'S ON TAP?

As the digital universe expands, so will storage capacity. Here is a recap of the technologies evolving to help storage keep pace with the growth of the digital universe.

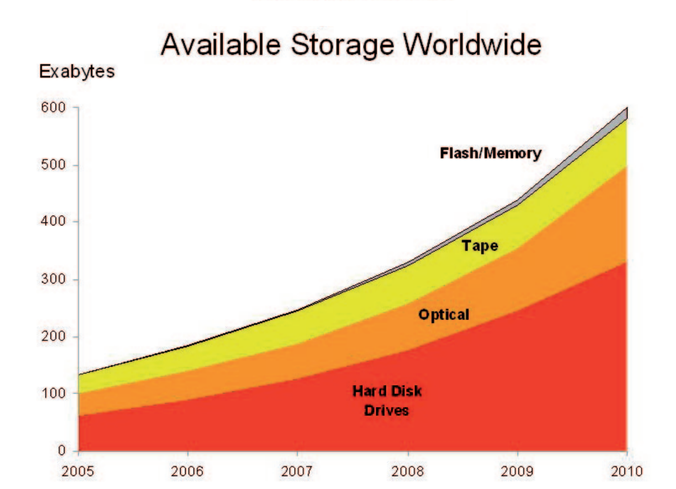
Hard disk drives continue to provide more storage capacity every year. In 2007, the first terabyte drive – 1,000 gigabytes! – will ship. Although we expect to see capacities continually increase, we will also see a parallel trend toward smaller disks. Some of the advanced technologies promising to take capacities to 2 terabytes and beyond include perpendicular recording (which packs more bits per inch on a disk platter than traditional recording), patterned media (a.k.a., nanobits, a new arrangement for storing bits on a platter), and heat-assisted magnetic recording (which reduces the amount of magnetism needed to store a bit).

Tape storage is the most prominently used back up and archive medium in large corporations. But, tape has its disadvantages and is being relegated to long-term archive and disaster recovery as alternative solutions emerge. Today's tape cartridges range in uncompressed capacity from less than 1GB per single cartridge to 500GB. Improvements in tape cartridge capacity come in one of three ways: Increasing linear/track density, thinning the media (so that more linear square feet can be wound on a cartridge), or increasing the tape width. We expect tape cartridge density to increase around 40% per year.

Optical storage, in the form of compact discs (CDs) and digital versatile discs (DVDs), is ubiquitous in today's society. While CDs and DVDs are mostly used for distributing content (e.g., movies and software), they can and are used for information archiving. One promising next-generation optical storage technology is holographic storage, which promises very stable long term storage in very dense packages. The first commercial holographic products should be available this year.

Nonvolatile flash memory – also called thumb drives, memory sticks, and USB memory – has seen rapid price declines, which have enabled its use in devices from cell phones and handheld games to industrial electronics and network components. As prices continue to fall, flash will see more use in portable electronics and even in solid state memory for laptop PCs. Ultimately flash may enable us to carry our own PC system profiles with us, so we can boot up on any PC and have all our data and applications ready for use.

Figure 10



Source: IDC, 2007

matter and one we don't address here. The percentage of available storage on hard disk drives will actually grow to more than half of total available storage in 2010.

Despite the growth of digital information associated with user created content and consumer electronics, the share of available storage tied to organizational information remains remarkably stable.

Driving the growth of organizational storage are a number of factors:

- The growth of networked communications, such as email and, increasingly, voice over IP, that require archiving.
- The growth of corporate data tied to increasing levels of automation and mission-critical applications, such as supply chain management, collaboration, product design, and customer self service.
- Regulations mandating new archiving and privacy protection rules.
- Industry specific applications, such as security imaging, RFID and sensors, Internet commerce, and medical records and imaging.
- Increased computerization of small businesses.
- The need for organizations to facilitate the exchange, distribution, and protection of consumer-driven information.

HOW WILL WE DEAL WITH ALL THIS CONTENT?

Managing the digital universe is not simply a matter of having enough storage capacity to store what we want. Those economics seem to work out over time, and, in fact, are linked. The growth of camera phones can proceed, in part, because of the available on-board storage. The growth of the volume server market can proceed apace because storage continually gets cheaper. We build the applications to fill the storage we have available, and we build the storage to fit the applications and data we have.

But will we be able to do useful things with the information we have? Or will all these exabytes become the equivalent of a trillion old photographs kept in an electronic shoebox?

Perhaps a little of both.

The cost of not responding to the avalanche of information can add up, yet not be immediately visible to CEOs and CFOs. In surveys of U.S. companies, we have found that information workers spend 14.5 hours per week reading and answering e-mail, 13.3 hours creating documents, 9.6 hours searching for information, and 9.5 hours analyzing information.

We estimate that an organization employing 1,000 knowledge workers loses $5.7 million annually just in time wasted having to reformat information as they move among applications. Not finding information costs that same organization an additional $5.3 million a year.

Adopting a comprehensive and disciplined approach to managing information and understanding its value is a key to reducing the hidden – and not so hidden – costs associated with the information explosion.

DETERMINING THE VALUE OF INFORMATION: INFORMATION LIFECYCLE MANAGEMENT

Not all digital bits are created equal. For consumers, family photos are probably worth more than last year's record of sent emails. Most people know what they would grab first if they had to vacate their homes in a hurry.

In organizations, where there are a lot more pieces of information, often buried in thousands of business applications

HOW LONG WILL OUR DIGITAL ARCHIVES LAST?

One paradox of the digital universe is this: Even as our ability to store digital bits increases, our ability to store them over time decreases.

We can read cuneiforms of clay tablets thousands of years old, scrolls and books over a thousand years old, and microfilm that is a hundred years old. But can we read a 8-track tape from 30 years ago, a floppy disk from 20 years ago, or a VHS tape from 10 years ago?

The life-span of digital recording media is nowhere near as long as stone or paper – the media degrades and the playback mechanisms become obsolete. The design life of a low cost hard drive is 5 years; the usable lifespan of magnetic tape has been estimated to be as little as 10 years,[viii] and the life expectancy of CDs and DVDs may be as little as 20 years.[ix]

In short, the life of stored data follows two conflicting curves: one where capacities go up and one where longevity goes down.

For the moment the solution recommended to digital archivists by the National Media Lab is to transcribe digital records to new media every 10-20 years – a tough assignment for all but the well-organized.

But long-term solutions are on the way, being spear-headed by the Storage Networking Industry Association (SNIA) via its 100-year archive initiative, as well as its work with leading storage companies on Extensible Access Method (XAM). Its implementations are expected to be several types of interfaces between applications and storage systems that coordinate metadata to achieve interoperability, storage transparency, and automation for ILM-based practices, long term records retention, and information assurance (security).

and with many more internal customers, determining which digital bits are worth more than which other digital bits is only now becoming an emergent science.

The process of determining the value of information is generally referred to as "information lifecycle management," or ILM. The Storage Networking Industry Association (SNIA) defines ILM as:

> *The policies, processes, practices, services and tools used to align the business value of information with the most appropriate and cost-effective infrastructure from the time information is created through its final disposition. Information is aligned with business requirements through management policies and service levels associated with applications, metadata and data.*

The ILM concept is simple to conceive. It basically means assigning a value to information, changing that value over time, storing the information according to its value, and deleting it when the time comes.

In practice, it's more difficult. Every user in an organization thinks his or her information is important, and data that seem worthless one day may all of a sudden become invaluable when auditors or lawyers request them. Value can not always be determined by the age of the data.

So ILM initiatives usually proceed in stages, starting with the development of a tiered storage architecture. Mission-critical transactional data might be stored on high performance disk systems attached to servers; less critical data, like year-old inventory history, might be stored on a storage area network populated with slower, less costly drives; the least critical information, such as document back-ups from organizational PCs, might be stored on tape.

ILM tools and solutions can include hardware, software, and consulting services that help companies classify information by value and manage it by class. Just the software market for managing archiving and hierarchical storage management is expected to double in size from 2006 to 2010 to $1.7 billion worldwide.

How much of today's information is "classified," or ranked according to value? IDC estimates that within organizations it is still less than 10%. About a quarter of organizational information lies outside the data center, and as much as 30% of corporate information sits in small businesses. But given the brisk growth in software and services around information management, IDC expects the amount of classified data to grow better than 50% a year.

But so will the total amount of data, so the percentage of data in the digital universe that is classified will grow more slowly.

THE UNSTRUCTURED DATA PROBLEM

Over 95% of the digital universe is "unstructured data" – meaning its content can't be truly represented by its location in a computer record, such as name, address, or date of last transaction. Digital images, voice packets, and the musical notes in an MP3 file would be considered unstructured data. In organizations, unstructured data accounts for more than 80% of all information.

There may be information about the content, such as when it was captured – e.g., the time stamp on a camcorder clip – or its compression scheme, address from which it was sent or received if indeed it was, or file size. But that information, or "metadata," is generally not enough to determine what is

STORING A LIFE'S WORTH OF INFORMATION

While science fiction writers have long imagined systems for storing and playing back all the events of our lives, at least one industry luminary is trying to do it for real.

In 2000, Gordon Bell, for many years the top technologist at Digital Equipment Corporation and now a senior researcher at Microsoft, began attempting to store all the information he creates and captures[x]. The project originally stored encoded archival material, such as books he read, music he listened to, or documents he created on his PC. It then evolved to capturing audio recordings of conversations, phone calls, web pages accessed, medical information, and even pictures captured by a camera that automatically takes pictures when its sensors indicate that the user might want a photograph.

The original plan was to test the hypothesis that an individual could store a lifetime's worth of information on a single terabyte drive, which, if compressed and excluding prerecorded video (movies or TV shows he watched) still seems possible. The experiment is also being used to work on the software and database technology to manage the storage and retrieval of the accumulated information.

By 2007, Bell had accumulated 300,000 records taking about 150 gigabytes of storage, so he probably could fit the information on a 1-terabyte disk were one available. However, in one experiment where TV programs he watched were recorded, he quickly ran up 2 terabytes of storage. So the one terabyte capacity is considered reasonable for text-audio recording at 20th century resolutions, but not full video.

In his experiment, Bell mimicked one of the trends we forecast for the digital universe. In 2000 he was shooting digital camera pictures at 2 MB per image; when he got a new camera in 2005 the images swelled to 5 MB. Along the way his email files got bigger as his attachments got bigger.

So let's see, at one terabyte per person, if everyone on the planet recorded everything Gordon Bell did, that would mean we'd need 620 exabytes of storage – about 30 times what's available today. Hmm.

DEDUPLICATION: LESS IS MORE

Duplicated information, while difficult to measure, has always been a major driver of capacity needs for storage. Beyond the vast distribution of various applications that are duplicated on servers and PCs worldwide, there is also the notion of backing up duplicate copies of information.

Deduplication (a.k.a., record linkage and single instancing) is a technology that can identify and manage the removal of duplicate data by linking records that refer to the same data within structured, as well as unstructured data environments. Think of not backing up only documents that have changed since the last backup, but backing up only the words and phrases in the document that have changed.

Reducing this redundancy within backed up information can yield efficiencies as much as 20:1 over traditional information backup methods today and potentially greater in the future.

actually contained in a unit of information without some human or automated intervention.

IDC believes that over time it will become easier to deal with unstructured data as (1) more and more metadata is added to unstructured data, (2) structure is added to unstructured data, and (3) access systems provide structured views of both structured and unstructured data.

Some of the current research directions we have seen supporting this belief include:

- Techniques for automatically classifying unstructured databased on examining the content itself, and then making it available to computer applications. The combined structured and unstructured data can then be displayed as the output of database queries.
- Methods for adding "structure" to unstructured content by examining the words or images in context, using "fuzzy" matching techniques, employing indexing engines, and so on.
- Tools to optimize multimedia searches, e.g., research on techniques to match images on the Worldwide Web to images on a handheld device.
- Tools for searching images, although today most of these require the user or a software program to "tag" the image or provide key words; most image search today is based on text analytics on information about the images. Researchers, however, are working on getting information about the image by examining the image or video stream itself for recognizable artifacts.
- Methods of searching audio files by the spoken word, rather than on textual metadata, which will help as audio information delivery, such as podcasts, grows.
- Work by the Worldwide Web Consortium, under the leadership of Tim Berners-Lee, inventor of the Web, on creating the "semantic Web," which would use automated tagging of data to allow searching inside XML documents.

An example of this type of technique can be seen in public safety applications in the Dublin, Ireland, airport, Milan, Italy, metro, and Dover, England, port. Here surveillance cameras are linked into information systems, and software using "fuzzy matching" techniques segments the scene and matches images to patterns it knows of suspicious behavior, such as someone leaving an unattended suitcase. The match doesn't need to be exact – just enough to call attention to the image.

COMPLIANCE: AN ORGANIZING FORCE

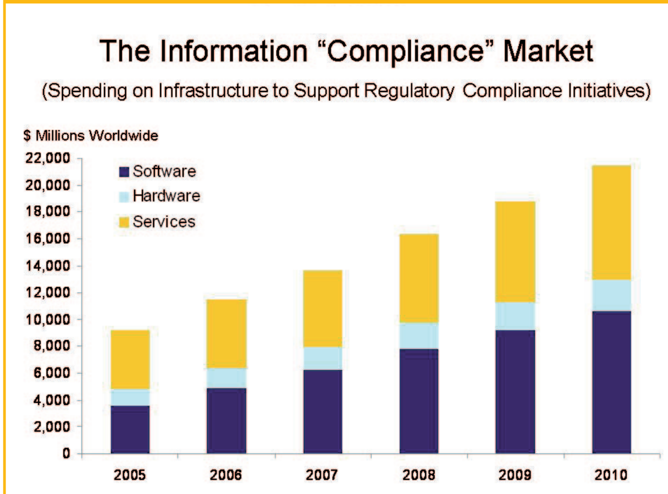
Sarbanes-Oxley, HIPAA[xi], SEC rule 17a-4[xii], FDA Rule 21 CFR Part 11[xiii], Basel II[xiv], and hundreds of other acronyms and initials refer to the collection of government and trade group rules that, together, the industry calls "compliance."

In meeting these rules, companies are forced to deal with new aspects of the expanding digital universe – since the rules set standards for record keeping, records retention, information security, and privacy protection, among other things. New rules for things like legal discovery of documents, called "e-discovery," are driving companies to formalize new records management policies, develop archiving standards, and institute policy changes and employee training.

How much of the digital universe is subject to compliance rules and standards? It's not easy to tell, but our best estimate is, today, about 20%. Records and files about company operations, employees, and customers, surveillance camera images that might be called in discovery, recorded phone calls and emails, records of online transactions, and so on. Given the current rate of growth of digital surveillance cameras alone, this percentage should grow steadily.

As a result, compliance is big business. Spending on just the hardware, software, and computer services to develop an IT infrastructure to support compliance initiatives is expected to double from 2006 to 2010 to $21.4 billion worldwide (Figure 11).

Figure 11



Source: IDC, 2007

The email archiving market alone is expected to be a billion dollar software and hosted services market by 2010, up from $471 million in 2006.

And this is spending that excludes money for consultants, lawyers, internal training, and the other non-IT costs of meeting the new rules.

There are benefits to companies that institute solid compliance strategies that go beyond simply fulfilling legal requirements. The compliance infrastructure of which IDC talks includes the convergence of content, collaboration, storage, security, and system and network management as applied to supporting compliance initiatives. Firms that get this part of their IT operations organized are generally going to be better at governing the rest of their business.

But the expansion of the digital universe will create challenges for even the most well-organized organizations. The new rules don't distinguish between structured and unstructured data, and the new information types streaming into and out of organizations will need to be folded into this compliance infrastructure. These include, in addition to emails, database records, and documents:

- Instant messages in organizations – which will be sent and received from 250 million IM accounts by 2010, including consumer accounts from which business IMs are sent;
- Voice over IP phone calls, which will increasingly be integrated over organizational IT networks;
- Web conferences, including both audio and video and both host-based and premises-based, and which will increasingly be an embedded function in other business applications;
- Multimedia publications, including podcasts and videocasts;
- Surveillance and security camera images, the digital versions of which are expected to grow more than 10-fold in the next 4 years.

IDC believes that compliance and risk management will be a key driving force for spending on software, services, and storage for years to come.

SECURITY: THE DARK SIDE OF THE DIGITAL UNIVERSE

The first highly publicized security breach of the Internet was a 99-line "worm" created by Robert Morris, a 23-year old student at Cornell, in 1998. The Internet Worm, as it came to be called, infected an estimated 6,000 computers in a matter of hours, slowing them down to the point where they had to be shut down and disinfected. The attack created headlines in the world press and earned its creator a fine, community service, and three years' probation. This is considered the first "worm."

This may have been the first worm, but it certainly wasn't the last.

Security threats in the digital universe have migrated from hacker pranks and disgruntled worker vandalism to sophisticated identity thefts that involve tricking users to come to fake PayPal sites, denial-of-service attacks for ransom, and creation of worms that can create or attack other worms.

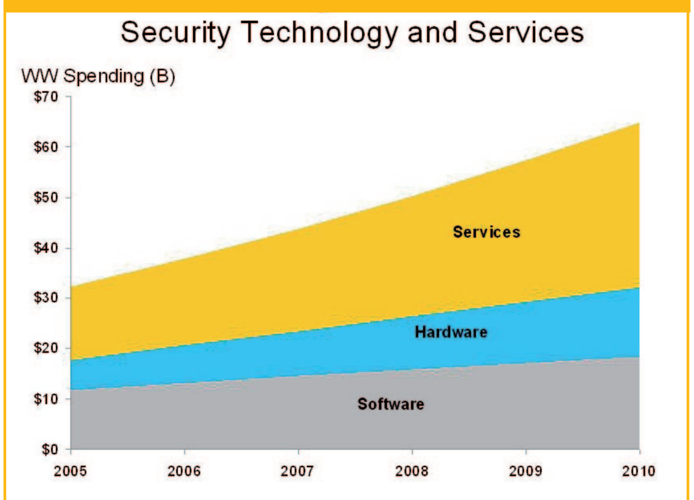
(In the past, security breaches were often not publicized. But as attacks on information security have become more sophisticated – and press reports on lost laptops with personnel records, theft of credit card and customer records by hackers, and other stories of compromised data more commonplace – the issue has entered the public eye. California, for instance, now requires companies and government agencies to disclose breaches of information confidentiality.)

In keeping with the increased threats and in addition to hardening the peripheral security, organizations are moving to secure access and identity, layer digital rights management on the information itself, increase security attributes of the systems including storage systems and encrypt backup tapes and other media that goes off site. To ensure compliance with their security policies and related regulations, organizations are consolidating, storing and analyzing substantial amounts of log data related to security applications and infrastructure elements.

Spending on security-specific software is already nearly $40 billion a year; by 2010 it will be $65 billion, or close to 5% of total IT spending. Add the software, hardware, and networks needed to support those security products and you are up over 10% of IT spending (Figure 12).

The growth of the digital universe almost dictates that the share of IT spending devoted to information security and privacy protection will have to go up. Billions of new users on the internet, increasingly sophisticated attackers, and many more networked signals almost ensure that security issues will get worse, not better. IDC estimates that today close to 30% of the bits in the digital universe are potentially subject to security applications; by 2010 that proportion will be closer to 40%.

Figure 12



Source: IDC, 2007

Thankfully, the pressure on organizations to organize their information more rationally to support compliance initiatives will dovetail nicely with the need to increase information security and privacy protection. This pressure will, in turn, drive the development of new organizational information infrastructures.

WHERE DO WE GO FROM HERE?

The digital universe is not only expanding, it is changing character and, along the way, changing the expectations and habits of people who use and depend on information. As we have seen:

- There is a comingling of user created and organization-managed information – from employees sending personal emails from corporate laptops to consumers listening to corporate podcasts on their own players.
- Most of the digital universe will remain unstructured – meaning tools and techniques will be required to add structure to this content to improve search, discovery, management, security, and storage.
- Most of the bits in the digital universe will travel over networks, including the Internet, new IP-based telephone networks, and broadcast networks.
- By 2007 there will be more information created, captured, and replicated than the capacity to store all of it.

These characteristics of the digital universe will affect us in profound ways. For consumers and citizens it means the continuation of the digital onslaught that really began in the 1990s with the personal computer, the Internet, and the cell phone.

Now we are adding more and more digital devices to the equation, from the new crop of video games and automobile GPS systems to programmable implants, RFID chips for marathon runners, and Bluetooth sunglasses.

The societal impact of the digital universe seems fairly straightforward – digital information literacy will be an increasingly important life skill. At the same time, the cost to access the bits in the digital universe will drop to the cost of a cell phone.

For organizations, the impact of the expansion of the digital universe is clear in the broad outline. We will need more storage, and more intelligent storage. We will need better management of the information we create, replicate, and store, and we will need to meet demands of both legal regulations and the competitive dynamics of our industries. More specifically:

- The growth of the digital universe means a sea change in the way organizations deal with customers, suppliers, and employees. IDC predicts that just the number of electronic commerce connections among companies and their customers will increase 100-fold in five years. This will drive new engagement strategies and techniques, from creating and managing blogs to mining customer data and integrating disparate applications, in ways we don't utilize today.
- IT managers will see the span of their domains expand rapidly – as VoIP phones come onto corporate network, building automation and security migrates to IP networks, surveillance goes digital, and RFID and sensor networks proliferate.
- IT and storage managers will need to step up their ILM and compliance infrastructure initiatives if they are to keep up with the quantity, character, and source of the digital information growing within organizations. Given the growth of the digital universe this feels like a race against time.
- Information security and privacy protection will rise to a boardroom concern as organizations and their customers become increasingly tied together with real-time connections. This will require not just the implementation of new technologies – from multimedia encryption and digital watermarking to advanced biometrics – but also new training, policies, and procedures. Physical security and information security will merge.
- The community of those with access to corporate data will become more diffuse – as workers become more mobile, as companies increasingly implement customer self service, and as globalization diversifies customer and partner relationships and elongates supply chains.

IDC's vision of "Dynamic IT" to support dynamic organizations sees the emergence of resource pooling via technologies such as virtualization and service-oriented-software decoupling rigid connections among computers, storage, data, and individual applications, where information today is trapped in narrow, hard to access silos.

Dynamic IT frees information from the underlying traditional IT infrastructure that stores, manages, and secures it. Thus, information can become the design center for advanced information infrastructures. Organizations today are beginning to re-architect their infrastructures to make them more dynamic and information-centric.

Along with the more flexible, dynamic, and service oriented infrastructure, by the way, will come more flexible, dynamic, and service oriented IT organizations.

So, today, we see enlightened organizations taking steps to keep up with the demands of an expanding digital universe by:

- Creating more service oriented IT organization by embedding staff within business units, developing service agreements with internal customers, and using business metrics to set IT performance goals.
- Establishing service level objectives for various IT functions, including information storage, management, security.
- Developing organization-wide policies for information security, records and email retention, privacy protection, and data access – and providing continual training and systems support to facilitate the policies.
- Refreshing the IT infrastructure with the new Dynamic IT tools, such as server and storage virtualization, database federation, business-rule base application development, automatic data center provisioning, and business analytics.
- Deploying advanced technologies for information capture, search and discovery, and data classification and tagging in pilot projects; and proselytizing the concepts of information lifecycle management.
- Establishing a Chief Security Office and elevating it from an IT function to a corporate function.

Stay tuned. There is still an information universe beyond the digital universe. The use of paper in the world is still increasing (nearly 5% in the last five years)[xv] . But we have clearly hit a threshold – where the digital universe is now pervasive enough to be a major locale for commerce, education, social interaction, and entertainment. It's only going to get bigger.

METHODOLOGY AND KEY ASSUMPTIONS

Our basic approach of sizing the expanding digital universe was to:

- Develop a forecast for the installed base of 49 classes of devices or applications that could capture or create digital information.
- Estimate how many units of information – files, images, songs, minutes of video, phone calls, packets of information -- were created in a year
- Convert these units to megabytes using assumptions about resolutions, digital conversion rates, and usage patterns.
- Estimate the number of times a unit of information is replicated, either to share or store. The latter is a small number, for example the number of spreadsheets shared, or a large number, such as the number of movies put into DVD or songs uploaded onto a peer-to-peer network.

Much of this information is part of IDC's ongoing research. For instance, we have published research on shipments and installed base for almost all of the devices we forecasted, we have the number of megabytes created by all digital cameras, camcorders, and camera phones, the number of megabytes of email traffic, and the average number of original documents created on PCs.

For devices like the PC and servers, we analyzed the information creation, capture or replication based on application or workload. For instance, IDC has amassed detailed information on server workloads as a result of decades of studying the server market and surveying users.

Since most of IDC's forecasts also have geographic splits, we were able to build our forecasts of the digital universe by region and aggregate to worldwide.

Below is a list of the kinds of devices or information categories we examined.

INFORMATION CREATION DEVICES

Image Capture/Creation

- High End Cameras
- Digital Cameras
- Camcorders
- Camera phones
- Webcams
- Surveillance
- Scanners
- Multifunction Peripherals
- OCR
- Bar Code Readers
- Medical Imaging
- Digital TV
- Digitized Movies & Video
- Special Effects
- Graphics Workstations

Digital Voice Capture

- Landline Telephony
- Voice over IP
- Mobile Phones

Data Creation

- PC applications
  - database
  - Office Applications
  - Email
  - Video/teleconference
  - IM
  - Other
- Smart Handhelds
- Terminals, ATMs, Kiosks,
- Specialized Computers
- Industrial machines/cars/toys
- RFID
- Sensors
- Smart Cards
- Video games
- MP3 players
- SMS
- GPS
- Server Workloads
  - Business Processing
  - Decision Support
  - Collaborative
  - Application Development
  - IT Infrastructure
  - Web Infrastructure
  - Technical
  - Other

Data Storage

- HDD
- Optical
- Tape
- NV Flash Memory
- Memory

AVAILABLE STORAGE

IDC routinely tracks the terabytes of disk storage shipped each year by region, drive type, and application type (e.g., PC, digital camera, storage array, etc.). We also track shipments of tape and optical drives.

To develop available storage on hard drives, IDC storage analysts estimated storage utilization on capacity shipped in previous years and added that to the current year shipments.

For optical and nonvolatile flash memory, we developed installed capacity ratios per drive and algorithms for capacity utilization and over-writing. In optical we found there was much more pre-recorded and write-once storage than storage that was over-written by users.

SOME KEY ASSUMPTIONS

Most assumptions about capacities, resolutions, compression or replication are embedded in the original source IDC forecasts. These are not repeated here.

But there were some assumptions that were material to the output:

- Images: We assumed that most images were captured or replicated in a compressed format, e.g., JPEG 100, rather than in raw format. This lowered overall exabytes.
- Digital TV: We counted creation as the creation of the content shown on the TV and its broadcast as replication. We only counted broadcasts that were seen, thus the growth in Digital TV exabytes tracks the deployment of digital TVs.
- Voice capture: Although many calls on the traditional voice network originate as analog signals, we assume that at somewhere in the network they are sampled and changed to digital pulses.
- Voice replication: Because we had no publicly-available data on the number of phone calls that might be recorded and stored by either phone companies or governments, we held replication to zero on all but VoIP calls. We did, however, estimate a small percentage of the voice traffic as data about the calls for billing and tracking.
- Computers: While computers store and transport many of the files, images, voice packets, and songs in the digital universe, we estimated the amount of information associated with them based on their end devices. This is one of the reasons the "data" portion seems small compared with the image portion. We did not count an image replicated from a digital camera to a PC as one that was "captured" by the PC but as a replication from the camera. We did this to avoid over counting (a replication as another device's capture).
- Music: Our estimate was based on assessing the total number of new songs created, which we assumed were created in a large file-format CD for distribution. Songs in MP3 format were considered replication. We estimated the number of legal song sales (CD and Web distribution) and added a conservative estimate of songs illegally distributed. It is quite possible that we were too conservative in our estimate of illegally shared songs over peer-to-peer networks.
- Sensors and RFID tags: We assumed a steady increase in the frequency with which the signals were read – from, for instance, weekly to multiple times a day in the case of an RFID tag. Despite the number of tags and sensors expected to be in use in the upcoming years, this did not affect total exabytes very much because the signals tend to be quite small. It did affect the total number of information "units" charted.

QUALITATIVE ESTIMATES

In some cases we developed estimates of information categories – e.g., structured versus unstructured content – by estimating the percent of a device type category that would apply (e.g., images as "unstructured.") These we then added up to develop percentages of the digital universe. These include our estimates of structured and unstructured data, organizational and consumer data, security or compliance intense data, and data that would would involve organization "touch." In this regard they are more subjective than the sizing of the digital universe based on device output.

LEVELS OF AGGREGATION

While we developed our forecast of the digital universe, we elected to aggregate our results into the categories shown to simplify our story and to protect the investment of the clients who have paid for the underlying foundation research. Where appropriate, we used IDC proprietary data to make our points.

The bibliography of supporting IDC reports gives you an idea of the extent of that underlying research.

BIBLIOGRAPHY

- Worldwide Plasma Display Panel 2006-2010 Forecast (IDC #201988, June 2006)
- Worldwide LCD TV 2006-2010 Forecast (IDC #201593, May 2006)
- Worldwide DVR 2006-2010 Forecast (IDC #204071, October 2006)
- Worldwide Digital Television Semiconductor 2006-2010 Forecast (IDC #203201, September 2006)
- Asia/Pacific (Excluding Japan) Digital Set-Top Box 2006-2010 Forecast (IDC #AP654108N, October 2006)
- Worldwide and U.S. Digital Pay TV Set-Top Box 2006-2010 Forecast (IDC #204338, November 2006)
- U.S. Digital Cable, Satellite and Telco TV Subscriber 2005-2009 Forecast (IDC #34628, December 2005)
- Western Europe Digital TV Technologies Forecast 2005-2010 (IDC #KD04N, November 2006)
- 2006 U.S. Consumer Digital Imaging Survey (IDC #203900, October 2006)
- Worldwide Digital Image 2006-2010 Forecast: The Image Capture and Share Bible (IDC #204651, December 2006)
- Worldwide Consumer Video Content and Archive 2006-2010 Forecast (IDC #204640, December 2006)
- Worldwide Digital Camcorder 2006-2010 Forecast (IDC #203195, August 2006)
- Worldwide Digital Still Camera 2006-2010 Forecast Update (IDC #203675, September 2006)
- Worldwide PC Camera 2006-2010 Forecast (IDC #34962, March 2006)
- Worldwide Camera Phone and Videophone 2006-2010 Forecast (IDC #204456, December 2006)
- U.S. High-Speed Document Imaging Scanner 2006-2010 Forecast (IDC #203552, September 2006)
- Worldwide Flatbed Scanner 2006-2010 Forecast (IDC #203000, August 2006)
- 2006 U.S. Mobile Imaging Survey (IDC #203901, October 2006)
- Worldwide Content Management and Retrieval Services 2006-2010 Forecast (IDC #35076, March 2006)
- Worldwide Content Access Tools (Search and Discovery) 2006-2010 Forecast Update (IDC #203439, September 2006)
- Unified Access to Content and Data: Delivering a 360-Degree View of the Enterprise (IDC #34836, February 2006)
- Unified Access to Content and Data: Database and Data Integration Technologies Embrace Content (IDC #204843, December 2006)
- Unified Access to Information: Content Vendors Heed the Urge to Converge (IDC #202942, August 2006)
- U.S. Network Camera 2007-2011 Forecast (IDC #205402, January 2007)
- Worldwide High-Speed Document Imaging Scanner 2006-2010 Forecast (IDC #204929, January 2007)
- Portable Audio Device Survey Results: IDC's Consumer Markets Audio Telephone and Web Surveys (IDC #203090, September 2006)
- Worldwide and U.S. Portable Compressed Audio Player 2006-2010 Forecast (IDC #201325, April 2006)
- 2006 Compliance in Information Management Forum West Survey: End-User Attitudes and Investment Priorities (IDC #202615, July 2006)
- Worldwide Archive and Hierarchical Storage Management Software 2006-2010 Forecast Update (IDC #203150, August 2006)
- Worldwide IT Security Software, Hardware and Services 2006-2010: The Big Picture (IDC #204736, December 2006)
- Worldwide Email Usage 2005-2009 Forecast (IDC #34504, December 2005)
- Worldwide Email Archiving Applications 2006-2010 Forecast (IDC #203535, September 2006)
- Worldwide Compliance Infrastructure 2006-2010 Forecast (IDC #201961, June 2006)
- Worldwide Enterprise Instant Messaging Applications and Management Products 2006-2010 Forecast (IDC #203848, October 2006)

- Server Workloads 2005: Understanding the Applications Behind the Deployment (IDC #35069, March 2006)
- Worldwide Videogame Hardware and Software Forecast 2006-2010 (IDC #34683, January 2006)
- Worldwide Multifunction Peripheral 2006-2010 Forecast (IDC #204136, November 2006)
- Worldwide IP PBX and Hardware Desktop IP Phone 1H06 Vendor Shares (IDC #203949, October 2006)
- Worldwide IP PBX and IP Phones 2006-2010 Forecast Update (IDC #202531, July 2006)
- U.S. Residential VoIP Services 2006-2010 Forecast (IDC #201638, May 2006)
- U.S. Residential VoIP Handset 2006-2010 Forecast (IDC #204690, December 2006)
- Demystifying the Digital Oilfield (IDC #EI202344, July 2006)

ADDITIONAL DATA SOURCES

- IDC Worldwide Black Book
- IDC Worldwide Telecom Black Book
- IDC Worldwide PC Tracker
- IDC Worldwide Server Tracker
- IDC Worldwide Internet Commerce Market Model
- IDC Worldwide Smart Handheld Device Tracker
- IDC Worldwide Storage Tracker

FOOTNOTES

[i] Wired, December 2006, "The Rise of YouTube," p. 22.

[ii] Pirates of the Digital Millennium, Gantz and Rochester, FT Prentice Hall, 2005, p. 175.

[iii] IEEE Spectrum, July 2006, "Ring of Steel II," p. 12.

[iv] Computerworld, October 30, 2006, "Where Size is Opportunity," p. 22.

[v] Lyman, Peter and Hal R.Varian, "How Much Information," 2003. Retrieved from http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/

[vi] Prepared remarks before the Senate subcommittee on Science, Technology, and Space, October 28, 1999.

[vii] Evan Schuman, Ziff Davis Internet, October 13, 2004, "At Wal-Mart, World's Largest Retail Data Warehouse Gets Even Larger."

[viii] A 1995 article in Scientific American by Jeff Rothenberg entitled "Ensuring the Longevity of Digital Documents" engendered years of discussion and debate on just how long magnetic tape lasts.

[ix] NIST Special Publication 500-252.

[x] Gordon Bell and Jim Gemmell, "A Digital Life," Scientific American, March 2007, pp. 58-65.

[xi] Health Insurance Portability and Accountability Act.

[xii] Rules on electronic record-keeping for brokers and dealers.

[xiii] Rules on electronic record keeping by the U.S. Federal Drug Administration.

[xiv] Basel II refers to the International Convergence of Capital Measurement and Capital Standards - A Revised Framework accord between international banks on standards measuring the adequacy of a bank's capital. It sets standards for electronic record-keeping, among other things.

[xiv] EarthTrends.wri.org database - paper/paperboard products 2000-2004 and "No End to Paperwork," World Resources Institute (WRI).

About IDC

IDC is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 900 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 90 countries worldwide. For more than 43 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company. You can learn more about IDC by visiting www.idc.com.

IDC

Analyze the Future

NOTES

IDC
Analyze the Future

Global Headquarters:
5 Speen Street • Framingham, MA 01701
508.872.8200
www.idc.com