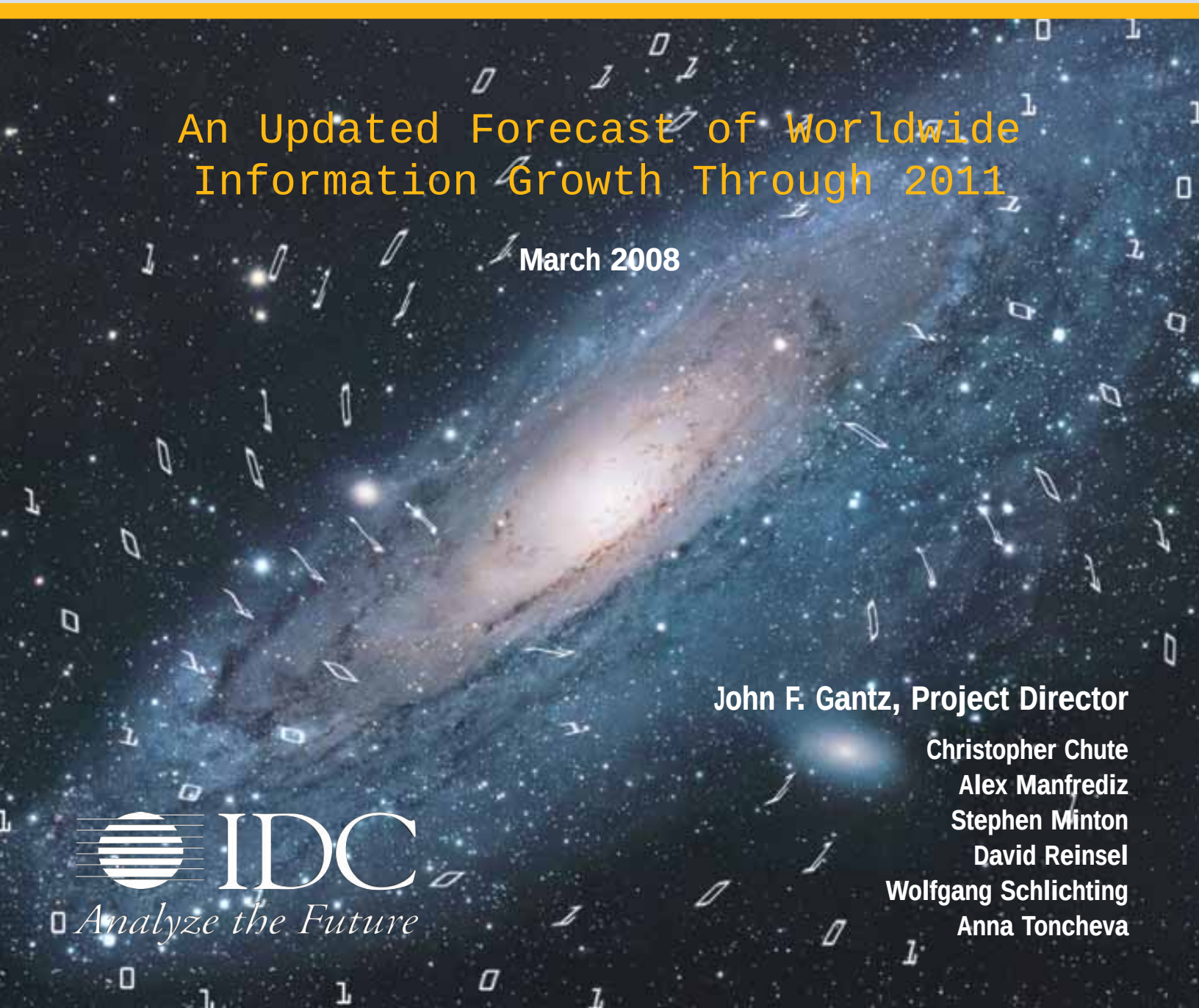
An IDC White Paper - sponsored by EMC

The Diverse and Exploding Digital Universe

An Updated Forecast of Worldwide Information Growth Through 2011

March 2008



John F. Gantz, Project Director

Christopher Chute
Alex Manfrediz
Stephen Minton
David Reinsel
Wolfgang Schlichting
Anna Toncheva

IDC
Analyze the Future

EXECUTIVE SUMMARY

This white paper, sponsored by EMC, is an update of IDC's inaugural forecast of the digital universe published in March 2007.[i] In this year's update we calibrate the size (bigger) and growth (faster) of the digital universe again, but we also explore some areas we only touched on last time. As before, we also seek to understand the implications for business, government, and society.

Some key findings are as follows:

- The digital universe in 2007 — at 2.25 x 10^{21} bits (281 exabytes or 281 billion gigabytes) — was 10% bigger than we thought. The resizing comes as a result of faster growth in cameras, digital TV shipments, and better understanding of information replication.
- By 2011, the digital universe will be 10 times the size it was in 2006.
- As forecast, the amount of information created, captured, or replicated exceeded available storage for the first time in 2007. Not all information created and transmitted gets stored, but by 2011, almost half of the digital universe will not have a permanent home.
- Fast-growing corners of the digital universe include those related to digital TV, surveillance cameras, Internet access in emerging countries, sensor-based applications, datacenters supporting "cloud computing," and social networks.
- The diversity of the digital universe can be seen in the variability of file sizes, from 6 gigabyte movies on DVD to 128-bit signals from RFID tags. Because of the growth of VoIP, sensors, and RFID, the number of electronic information "containers" — files, images, packets, tag contents — is growing 50% faster than the number of gigabytes. The information created in 2011 will be contained in more than 20 quadrillion — 20 million billion — of such containers, a tremendous management challenge for both businesses and consumers.
- Of that portion of the digital universe created by individuals, less than half can be accounted for by user activities — pictures taken, phone calls made, emails sent — while the rest constitutes a digital "shadow" — surveillance photos, Web search histories, financial transaction journals, mailing lists, and so on.
- The enterprise share of the digital universe is widely skewed by industry, having little relationship to GDP or IT spending. The finance industry, for instance, accounts for almost 20% of worldwide IT spending but only 6% of the digital universe. Meanwhile, media, entertainment, and communications industries will account for 10 times their share of the digital universe in 2011 as their share of worldwide gross economic output.
- The picture related to the source and governance of digital information remains intact: Approximately 70% of the digital universe is created by individuals, but enterprises are responsible for the security, privacy, reliability, and compliance of 85%.

To deal with this explosion of the digital universe in size and complexity, IT organizations will face three main imperatives:

One. They will need to transform their existing relationships with the business units. It will take all competent hands in an organization to deal with information creation, storage, management, security, retention, and disposal in an enterprise. Dealing with the digital universe is not a technical problem alone.

Two. They will need to spearhead the development of organizationwide policies for information governance: information security, information retention, data access, and compliance.

Three. They will need to rush new tools and standards into the organization, from storage optimization, unstructured data search, and database analytics to resource pooling (virtualization) and management and security tools. All will be required to make the information infrastructure as flexible, adaptable, and scalable as possible.

We have many of the tools in place — from Web 2.0 technologies and terabyte drives to unstructured data search software and the Semantic Web — to tame the digital universe. Done right, we can turn information growth into economic growth.

THE VISUAL UNIVERSE

Contemplating the digital universe is a little like contemplating Avogadro's number. It's big. Bigger than anything we can touch, feel, or see, and thus impossible to understand in context. For the purists, Avogadro's number — the number of carbon atoms in 12 grams[ii] — is 602,200,000,000,000,000,000,000, or 6.022×10^{23}. And no, the digital universe is not that big. In 2007, the number of "atoms" in the digital universe — the digital bits, or binary 1s and 0s created, captured, and replicated during the year — was less than a hundredth of Avogadro's number.

But the number of digital "atoms" in the digital universe is already bigger than the number of stars in the universe. And, because the digital universe is expanding by a factor of 10 every five years, in 15 years it *will* surpass Avogadro's number.

But the size and explosive growth of the digital universe are only two of its characteristics. Like our own physical universe, it is also incredibly diverse, has hotspots, and is subject to mysterious unseen forces. It seems to have its own laws of physics.

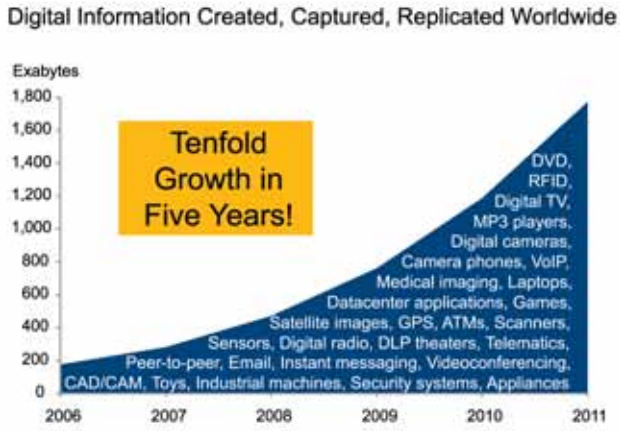
The IDC research shows that the digital universe — information that is either created, captured, or replicated in digital form — was 281 exabytes[iii] in 2007. In 2011, the amount of digital information produced in the year should equal nearly 1,800 exabytes, or 10 times that produced in 2006 (see Figure 1). The compound annual growth rate between now and 2011 is expected to be almost 60%.

The size of the digital universe in 2007 (and 2006) is bigger by 10% than we calculated last year, and the growth is slightly higher. This is a factor of faster-than-expected growth in higher-resolution digital cameras, surveillance cameras — especially in places like China and major urban centers — and digital TVs and of improved methodology for estimating replication.

The resolution of digital cameras and growth of surveillance cameras are important because the digital universe — at least in raw gigabytes — is predominantly visual: images, camcorder clips, digital TV signals, and surveillance streams.

The conversion from film to digital is practically over — last year the number of digital cameras and camera phones in the world surpassed 1 billion, and fewer than 10% of all still images captured were on film. Thus, when consumers buy higher-resolution cameras or camera phones, they have a measurable impact on the total gigabytes captured. A single image from a five-megapixel camera can be 40 megabytes uncompressed (1.2 megabytes compressed).

Figure 1



Source: IDC, 2008

In the surveillance world, the conversion to digital is in its infancy. Most cameras are still analog. But shipments of networked digital cameras are doubling every year. China is investing billions in video security systems for the Olympics and 2010 World's Fair, and it has a new "safe cities" policy that mandates security cameras for 660 cities and towns and 28,000 coal mines. New York City is rolling out a $90 million surveillance "veil" for Lower Manhattan. Police cars in many cities of the world now have mobile security cameras that can detect up to 200 license plates an hour.

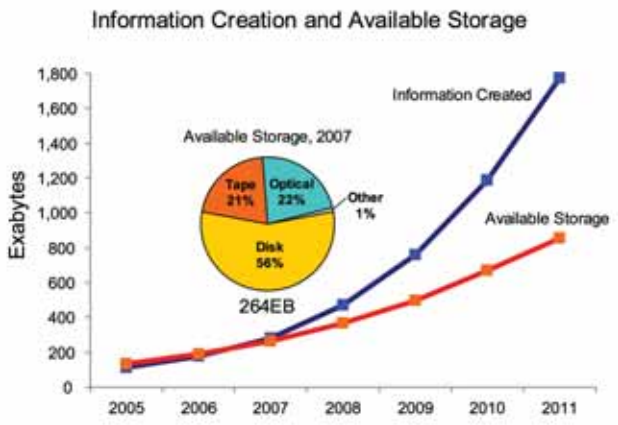
Finally, as analog TV systems in most countries of the world convert to digital in the next several years, digital bits will fly even more furiously. The number of digital TVs in the world doubled last year and should surpass 500 million by the end of 2011.

INFORMATION OVERLOAD GETS PHYSICAL

While the devices and applications that create or capture digital information are growing rapidly, so are the devices that store information. Information creation and available storage are the yin and yang of the digital universe. Cheaper storage allows us to take high-resolution photos on our cell phones, which in turn drives demand for more storage. Higher-capacity drives allow us to replicate more information, which drives growth of content. Yin, yang.

In 2007, according to our estimates, all the empty or usable space on hard drives, tapes, CDs, DVDs, and memory (volatile and nonvolatile) in the market equaled 264 exabytes — very close to the total amount of information created or captured (see Figure 2). From here on, the two numbers diverge.

Figure 2



Source: IDC, 2008

How to interpret this gap? Surely not all information created is important enough to store for any length of time, is it?

Correct. A good portion of the digital universe is transient — radio and TV broadcasts that are listened to but not recorded, voice call packets that are not needed when the call is over, images captured for a time then written over on a surveillance camera recorder.

But this is our first time in the situation where we couldn't store all the information we create even if we wanted to. This mismatch between creation and storage, plus increasing regulatory requirements for information retention, will put pressure on those responsible for developing strategies for storing, retaining, and purging information on a regular basis.

STORAGE BEATS EXPECTATIONS, TOO

When we put together our forecast of the digital universe last year, we estimated that 1,082 exabytes of storage would ship during the years 2007 through 2010. This time around, we've increased our estimates over the same time period by nearly 10%, or almost 90 exabytes.

Why? Three main reasons.

One. Protection of personal information. The segment of storage consumption most underestimated by IDC early in 2007 was that for personal data protection. Worldwide shipments of personal storage devices, a.k.a. external hard disk drives, exceeded all expectations in 2007. By 2011, personal storage devices are expected to consume more hard drive terabytes than all other segments except desktop PCs. As consumers generate more and more of the world's digital content, they are finally coming to understand the need to preserve their information heirlooms.

Two. Mobility. Increasingly, we carry our storage with us — in laptop PCs, mobile phones, iPods, PDAs, global positioning systems, games, and other computer electronics. Solid state storage in the form of flash memory is being driven into a broad spectrum of computing devices. And although flash represents a small percentage of overall storage capacity shipped — 1% in 2007 increasing to 5% in 2011 — our new forecast represents a cumulative 43% increase over the years 2007–2010 from our initial forecast last year.

Three. The side effect of storage on the go. Mobile phones, global positioning systems, PDAs, and other devices integrate local storage, but they also require access to networked storage across an increasingly connected world. This is one of the reasons enterprises are seeing their storage requirements increase 50% per year.

WRESTLING WITH DIVERSITY

There is another way to look at the digital universe besides in terms of gigabytes. What about the things *in* the digital universe? The equivalents of galaxies, stars, planets, asteroids, and specks of cosmic dust?

In our parlance, those celestial bodies would be images, video clips, TV shows, songs, voice packets, financial records, documents, sensor signals, emails, text messages, RFID tag transmissions, barcode scans, X-rays, satellite images, toll booth transponder pings, and the notes of "Happy Birthday" coming from singing greeting cards. Some of these things are big; some are small. An archived digital movie master kept at the National Academy of Arts and Sciences might be a terabyte. A DVD might be 5 gigabytes. An email a few kilobytes. An RFID signal only 128 bits.

THE DIGITAL UNIVERSE'S ENVIRONMENTAL FOOTPRINT

Tenfold growth of the digital universe in five years will have a measurable impact on the environment, in terms of both power consumed and electronic waste.

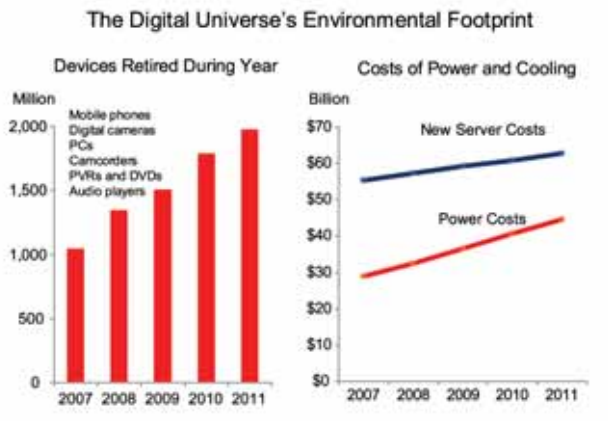
Electronic waste is already accumulating at more than 1 billion units a year — mostly mobile phones, but also personal digital electronics and PCs. The switch to digital TV will place a lot more analog TV sets and obsolete set-top boxes and DVDs on the waste pile, which will double by 2011.

Power consumption is harder to determine, especially as manufacturers develop power-saving chips and users install power-saving systems, including new cooling and air conditioning and new management systems (see Figure 3).

But in a study of server power and cooling costs conducted in 2006, IDC found that power and cooling costs are escalating rapidly as newer, denser servers come online. Power consumption that was 1kW per server rack in 2000 is now closer to 10kW. Customers building new datacenters are planning for 20kW per rack.

"Green IT" is a hot topic in IT circles today. With the expanding digital universe, discussion will have to turn to action quickly.

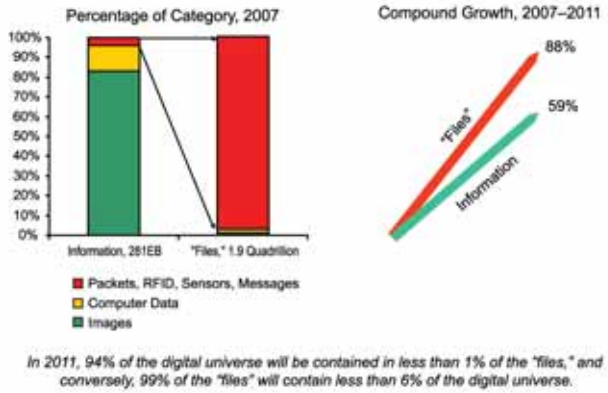
Figure 3



Source: IDC, 2008

In our physical universe, 98.5% of the known mass is invisible, composed of interstellar dust or what scientists call "dark matter."[iv] In the digital universe, we have our own form of dark matter — the tiny signals from sensors and RFID tags and the voice packets that make up less than 6% of the digital universe by gigabyte, but account for more than 99% of the "units," information "containers," or "files" in it (see Figure 4). The information created in 2011 will be contained in more than 20 quadrillion — 20 million billion — of these "files."

Figure 4



Source: IDC, 2008

This would not be an issue except that custodians of the digital universe — the technologists and datacenter managers working in enterprises, phone companies, ISPs, content and entertainment companies, and elsewhere — must keep track of all these little packets and signals. They must decide if, when, and how to store them, keep them secure, and adjust processes, sometimes in a split second, based on the content, however little, they contain.

The flip side of the problem occurs in the other 94% of the digital universe, where most of the content is opaque and unstructured within the file. Searching for meaning in the content of unstructured data like images, video clips, documents, and the numbers and characters in databases is the rocket science of the digital universe.

THE ENTERPRISE DILEMMA

We mentioned it last year, but a critical dilemma at the core of the digital universe remains. It's this:

> *While 70% or more of the digital universe is created, captured, or replicated by individuals — consumers and desk and information workers toiling far away from the datacenter — enterprises, at some point in time, have responsibility or liability for 85%.*

This responsibility includes information security, privacy protection, copyright protection, screening for obscenity, detecting fraud, reporting on and archiving the content, searching and retrieving, and disposal.

Examples abound. Consumers post video clips to YouTube, and Viacom sues Google for a billion dollars. Sixty million consumers trade pirated MP3s over peer-to-peer networks like Kazaa, LimeWire, and once Napster, and the record industry goes to war with ISPs to get consumer IP addresses. A video of a couple kissing at the metro in Shanghai appears on the Internet and results in a lawsuit for the Shanghai Metro Operations Company. Linden Lab launches a popular virtual world (Second Life) where visitors set up an economy based on virtual dollars — and gets sued for real dollars when a user who had invested in virtual land has his account terminated. The U.S. government makes USB memory sticks available to soldiers and later finds them being sold on the black market in Kabul — with sensitive data still on them.

IDC estimates that less than 5% of the digital universe actually emanates from datacenter servers, and only 35% emanates from the enterprise overall, mostly from workers at their desks, on the road, or working at home (see Figure 5).

This enterprise responsibility may be understood by corporate lawyers, investor relations staff, CEOs, and public relations specialists, but the technicians running the datacenter may not be well equipped to translate that understanding into datacenter policies, storage strategies, or information security practices. (See the Lessons for the Enterprise and Jumping to the Next Power of 10 sections later in this document.)

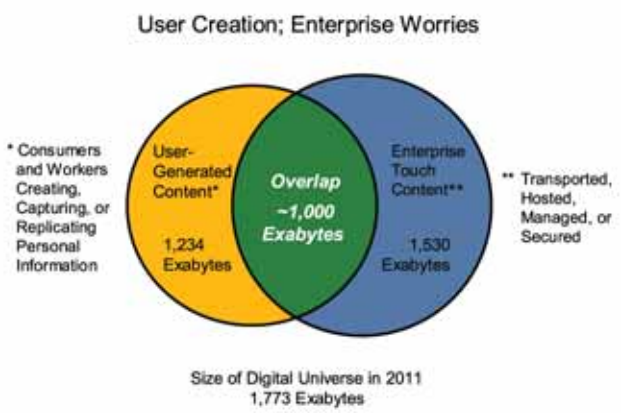
THE INDUSTRY KALEIDOSCOPE

With a little estimation, the digital universe can be divided into domains by industry. Do that, however, and you find a universe that does not resemble the world economy, the workforce, or the population. Instead, the digital universe follows rules of its own.

Take financial services, an industry synonymous with number crunching. Some of the most advanced computing takes place at brokerages, and some of the most meticulous record keeping occurs at insurance companies. Transactions involving trillions of dollars a day — equal to the world's annual gross economic output — are logged deep in the banking systems' computers. This is one of the reasons an industry that generated 6% of worldwide gross output buys 20% of the world's computers.

Yet for all this information processing, the financial services industry accounts for just 6% of the digital universe today and will fall to 3% by 2011. There is simply not enough imaging going on.

Figure 5



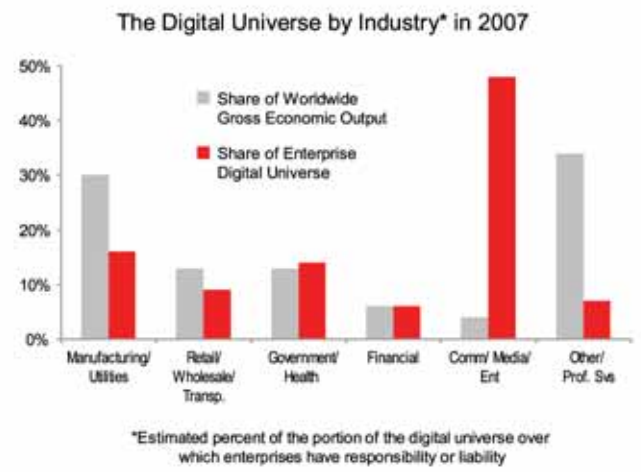
Source: IDC, 2008

On the other hand, there are the broadcast, media, and entertainment industries, which garner about 4% of the world's revenues but which already generate, manage, or otherwise oversee 50% of the digital universe. Within 10 years, when most countries are broadcasting digital TV and most movies are digital, that percentage will be even higher (see Figure 6).

Other industries have their own unique relationships with the digital universe:

- The manufacturing industry is rapidly deploying digital surveillance cameras on the one hand and sensor-based systems and RFID tracking on the other, not to mention using a lot of CAD/CAM and visualization.
- The retail/wholesale industry, here coupled with the transportation industry, is another major implementer of video surveillance and RFID tags. In addition, the rapid growth of customer information systems is swelling corporate

Figure 6



Source: IDC, 2008

databases. Wal-Mart, which now refreshes its customer databases hourly, adds a billion rows of new data an hour to a data warehouse that is 600 terabytes and growing.[v]

- The utility industry is talking about transforming the electricity distribution system into an "intelligent grid," with millions — perhaps billions — of sensors in the distribution system and at the meter level, broadband information transfer along power lines, and databases and active analytics to make system adjustments on the fly.
- Government and healthcare sectors are both heavily invested in imaging — surveillance and mapping in government, medical imaging and record archiving in healthcare. In healthcare, imaging databases are growing for two reasons: (1) growth of images per year (more patients, more scans) and (2) conversions of archived film images. A large hospital, like the Cleveland Clinic, might now have a petabyte-scale database of stored images and be adding to it as many as three terabytes a week.
- The oil and gas industry has been developing what's known as the "digital oilfield," where sensors monitoring activity at the point of exploration and the wellhead connect to information systems at headquarters and drive operational and exploration decisions in real time. Chevron has reported that it accumulates data at the rate of two terabytes a day. The raw geological data set for an oil field might be 200 terabytes.

There are also unique pockets of the digital universe worthy of note that can be tied to single entities. YouTube's 100 million video streams a day account for almost as much of the digital universe as all of medical imaging. The U.S. government's Center for Earth Resources Observation and Science has archives of three petabytes — mostly aerial photography and satellite images — and is growing at two terabytes a day. Library and archive digitization efforts, although small in the scheme of the entire digital universe, are steadily adding terabytes a day to the digital universe.

Then there is the new Large Hadron Collider (LHC) at CERN, the European Organization for Nuclear Research in Switzerland, which will go online this summer. When it runs an experiment, a system of sensors laid out in a plane the size of a swimming pool will gather data from four detectors at half a petabyte per second each, filter out most of the signals, then stream them at terabytes per second to an information grid. Just one experiment, the Compact Muon Solenoid (CMS), will receive incoming compressed data at 40 terabytes per second and store a megabyte per second.[vi] The experiment is expected to run 100 days a year, 24 hours a day. That's more than 300 exabytes of incoming data per year! The LHC will create a digital universe unto itself!

YOUR DIGITAL SHADOW

In last year's white paper, we reported on the effort by industry luminary Gordon Bell to digitally record his entire life. By the beginning of the year, he had accumulated 150 gigabytes of records, excluding TV shows or movies he watched.

How would that apply to us? To you?

In 2007, the digital universe contained 281,000,000,000 gigabytes, which works out to about 45 gigabytes per person on the planet.

Yet in 2007, when IDC developed the Personal Digital Footprint Calculator, launched this month,[vii] we discovered that only about half of the digital footprint would be related to individual actions — taking pictures, making VoIP phone calls, uploading videos to YouTube, downloading digital content, and so on.

We called the remainder "ambient" content. It is digital images of you on a surveillance camera and records in banking, brokerage, retail, airline, telephone, and medical databases. It is information about Web searches and general backup data. It is copies of hospital scans. In other words, it is information *about* you in cyberspace. Your digital shadow, if you will.

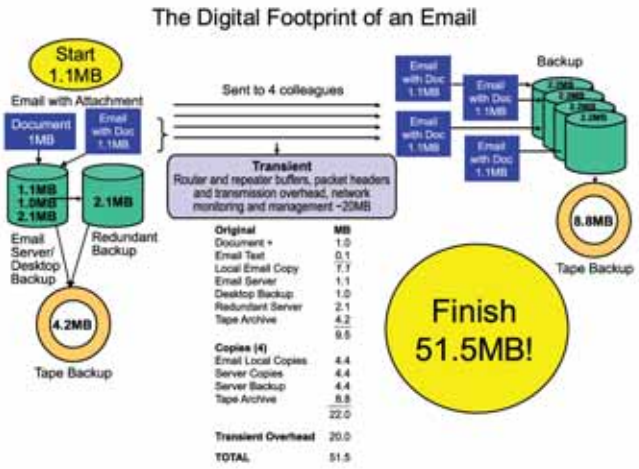
A DAY IN THE LIFE OF AN EMAIL

How the digital universe feeds upon itself can be seen in the digital footprint created by a single email sent to a team of four people, an example based on an email infrastructure similar to IDC's (see Figure 7).

The email itself is small, but with it is a 1MB attachment. If the email is sent to four people, wouldn't that mean that there are 5 x 1.1MB involved? The original and four copies?

No, unfortunately. To begin with, there is the document itself stored on the local machine, then the email that contains the document. In this infrastructure, copies of all emails are kept on the central email server, which, in order to keep the email system up and running, includes a redundant server. Desktop files, where the original document sits, are backed up daily to a server. The servers are then periodically backed up to tape and taken offsite. Our original 1.1MB email has a footprint eight times bigger than itself.

Figure 7



Source: IDC, 2008

Now add up the local and backed-up copies of the email sent to the four colleagues, and that footprint is 30 times larger than the original email.

Then there is all the temporary data created as the emails and backup systems send data back and forth across the local and wide area networks. In transmission, all manner of communications overhead is introduced: signaling data, packet addresses and headers, security codes, router caches, and management and tracking information. The estimate here is admittedly fuzzy, but it is within the order of magnitude.

There are techniques for deduplicating redundant emails and multiple copies of documents, but they aren't widely spread yet. In the meantime, a simple email can have a very long shadow.

Having a digital shadow is not necessarily bad. It allows Amazon to recommend new books to you, tells others they can trust you in an eBay transaction, and helps long-lost relatives find you. But it has a downside as well.

According to news reports, a citizen of Britain, with its estimated 5 million surveillance cameras, may expect to have his or her image captured 300 times a day.[viii] This has disturbed enough Brits that an underground group called Motorists Against Detection has begun burning traffic cameras.[ix] When Facebook began automatically tracking Web purchases by its members and sharing that data with others, users rebelled, and there are still issues about how difficult it is for inactive users to remove personal information from the site.[x] Taxi drivers in New York went on strike last September to protest plans for GPS vehicle tracking.[xi]

The idea of a digital shadow goes from curious or irritating to scary when you factor in the risk of identity theft. It was information about credit card purchases, including card and driver's license numbers, which was stolen from TJX by hackers working over a number of years and which exposed almost 50 million credit and debit cards to theft. According to the Ponemon Institute's *2007 Annual Study: Cost of a Data Breach*, it now costs companies almost $200 per customer record compromised in a security breach.

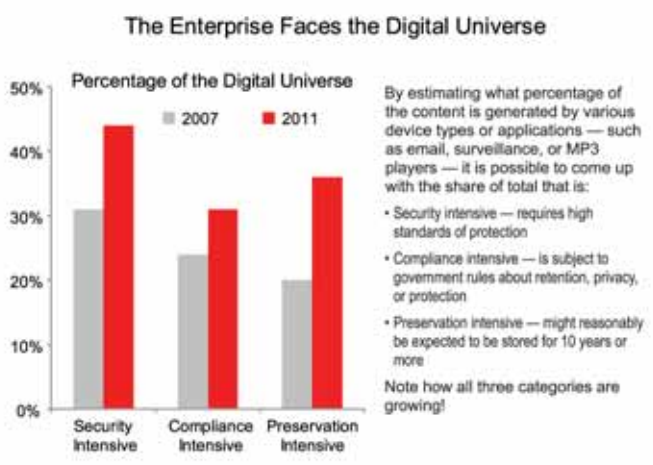
LESSONS FOR THE ENTERPRISE

Since the 2007 study was released, IDC has presented the results to thousands of CIOs and business executives in hundreds of conferences and meetings. We have learned some things from all this contact:

- The typical organization accepts the findings of the study, is already feeling the stress in storage management, and knows the stress will get worse; most are only very early in implementing information life-cycle management as their enterprisewide information management strategy.
- The typical CIO understands the security and privacy implications of the growing digital universe but is not sure how to get the rest of the company to understand them.
- Most CIOs and data professionals do not have a good handle on how the changing nature of the digital universe will change their relationship with end-user departments — which must be enlisted in the effort to classify, secure, and manage information coming from all sides into the organization.
- Few are ready to embrace the new data types — VoIP packets, surveillance videos, real-time sensor information — into their information management domain; few understand the potential impact on computing and information architecture.

Figure 8 shows a unique view of the digital universe by the degree to which the information in it might be subject to significant requirements for security; be subject to legal and compliance requirements such as ediscovery, HIPAA, or Sarbanes-Oxley; or be valuable enough to expect to store for 10 years or more.

Figure 8



Source: IDC, 2008

Not every camera phone image needs to be saved or archived, but account information and records on YouTube might be subject to ediscovery. Many emails emanating from inside corporate firewalls will be subject to some kinds of rules about retention or discovery. Search histories from search engines have been subpoenaed by the U.S. government.[xii]

The point of this exercise is not the numbers themselves but their order of magnitude. Because we plotted on a percent axis, the graphic doesn't show the raw growth in each category — which is much faster than the 10-times-in-five-years growth of the overall digital universe.

JUMPING TO THE NEXT POWER OF 10

The digital universe will be 10 times bigger in five years. What are we going to do about this?

As a society, our experience with the digital universe will unfold somewhat like a science-fiction novel. Within five years, there will be 2 billion people on the Internet and 3 billion mobile phone users. All will be interconnected; all will be creating and consuming content at an alarming rate. We can see fragments of the future today in the worlds of Second Life and Club Penguin, the stream of SMS messages to Twitter.com, the clinics in Beijing for exhausted Web addicts,[xiii] traffic control in Singapore, and sneakers that talk to officials of the New York Marathon.

For the custodians of the digital universe, however, the digital universe had better *not* unfold like a science-fiction novel. It needs to unfold like a dull, boring engineering text.

We can see the broad forces propelling the digital universe outward — mobility, interactivity, real-time information, user-created content, "compliance," new information form factors, and storage, storage, storage.

But to deal with so much change, IT organizations will face three main imperatives:

One. Transform their existing relationships with the business units. These are the groups that will classify information, set retention policies, deal with customers whose data the company holds, and face the public if data is lost, breached, compromised, or simply handled badly. Leading companies today are experimenting with embedding staff in line departments, charging for IT services based on business metrics, and routinely meeting with external customers.

Two. Spearhead the development of organizationwide policies for information security, information retention, data access,

WHAT WE COVERED LAST YEAR

This white paper is an update to last year's inaugural study (see **www.emc.com/digital_universe**) that refreshes the quantitative forecast of the digital universe and covers some new areas. It is meant as a companion to the original white paper. Some of the areas covered in more depth in last year's white paper are:

- Explanation of bits and bytes
- Analogs for the digital universe — its equivalent in books and elephants
- The growth of email, the Internet, and broadband communications
- The conversion of imaging, voice communications, and TV from analog to digital
- The digital universe by region
- Unstructured data
- "Compliance," the new rules driving the need to add structure and coherence to enterprise information
- Information life-cycle management
- Digital preservation
- Deduplication

and compliance. Extend these policies to business partners. Force the organization to mandate continual training in all these areas.

Three. Rush new tools and standards into the organization. Storage optimization, unstructured data search, database analytics, resource pooling (virtualization), and management and security tools — all will be needed to make the information infrastructure as flexible, adaptable, and scalable as possible.

Changes wrought by the digital universe will be swift and dramatic. But we have many of the tools in place — from Web 2.0 technologies and terabyte drives to unstructured data search software and the Semantic Web — to adjust to these changes.

The trick, and our challenge, will be to turn information growth into economic growth.

METHODOLOGY

Our basic approach of sizing the digital universe was to:

- Develop a forecast for the installed base of any of 30 or so classes of device or application that could capture or create digital information.
- Estimate how many units of information — files, images, songs, minutes of video, calls per capita, packets of information — were created in a year.
- Convert the units of information to megabytes using assumptions about resolutions, compression, and usage.
- Estimate the number of times a unit of information might be replicated, either to share or store.

Much of this information is part of IDC's ongoing research (see the Bibliography). Figure 9 provides a list of the kinds of devices or information categories we examined.

Figure 9

Devices and Applications Tracked

Image Capture/Creation
- High-end cameras
- Digital cameras
- Camcorders
- Camera phones
- Webcams
- Surveillance
- Scanners
- Multifunction peripherals
- OCR
- Barcode readers
- Medical imaging
- Digital TV
- Digitized movies and 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
- Server workloads
  - Business processing
  - Decision support
  - Collaborative
  - Application development
  - IT infrastructure
  - Web infrastructure
  - Technical
  - Other
- Terminals, ATMs, kiosks, specialized computers
- Industrial machines/cars/toys
- RFID
- Sensors
- Smart cards
- Videogames
- MP3 players
- SMS
- GPS

Data Storage
- HDD
- Optical
- Tape
- NV flash memory
- Memory

Source: IDC, 2008

AVAILABLE STORAGE

IDC routinely tracks the terabytes of disk storage shipped each year by region, media, and application. 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 device and algorithms for capacity utilization and overwriting. In optical, we found there was much more prerecorded storage than storage that was overwritten by users.

BIBLIOGRAPHY

- Worldwide Plasma Display Panel 2007–2011 Forecast and Analysis: Pure Price Play (IDC #206717, May 2007)
- Worldwide LCD TV 2007–2011 Forecast and Analysis: Exploding Units, Flatlined Revenue (IDC #206609, May 2007)
- Worldwide DVR 2007–2011 Forecast (IDC #210061, December 2007)
- Worldwide Digital Consumer Semiconductor 2007–2011 Forecast (IDC #210095, December 2007)
- Asia/Pacific (Excluding Japan) Digital Set-Top Box 2007–2011 Forecast and Analysis: 2006 Review (IDC #AP654111P, January 2008)
- 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 2007–2011 Forecast and Analysis (IDC #206623, May 2007)
- Western Europe Pay TV Digital Cable and Digital Satellite Technologies Forecast and Analysis, 2007–2011 (IDC #KD06P, November 2007)
- China Digital TV 2007–2011 Forecast and Analysis (IDC #CN656109P, May 2007)
- IDC's 2007 Consumer TV Survey, Part 1: Demographics and Current TV Preferences (IDC #209546, December 2007)
- 2007 U.S. Consumer Digital Imaging Survey (IDC #207516, July 2007)
- Worldwide Digital Image 2007–2011 Forecast: The Image Archive Bible (IDC #209873, December 2007)
- Worldwide Digital Image 2007–2011 Forecast: The Image Capture and Share Bible (IDC #209738, December 2007)
- Worldwide Digital Still Camera Memory Card Slot 2007–2011 Forecast (IDC #209316, November 2007)
- Worldwide Consumer Video Content and Archive 2007–2011 Forecast: The Video Bible (IDC #210035, December 2007)
- Worldwide Digital Camcorder 2007–2011 Forecast and 2006 Vendor Shares (IDC #208937, October 2007)
- Worldwide Digital Camcorder Storage 2007–2011 Forecast (IDC #209603, November 2007)
- Worldwide Digital Still Camera 2007–2011 Forecast Update (IDC #208141, August 2007)
- Worldwide PC Camera 2007–2011 Forecast and 2006 Vendor Shares (IDC #205559, February 2007)
- Worldwide Camera Phone and Videophone 2007–2011 Forecast (IDC #208561, September 2007)
- Worldwide Network Camera 2007–2011 Forecast (IDC #205402, January 2007)
- Worldwide High-Speed Document Imaging Scanner 2006–2010 Forecast (IDC #204929, January 2007)
- Worldwide Flatbed Scanner 2006–2010 Forecast (IDC #203000, August 2006)
- U.S. Flatbed Scanner 2007–2011 Forecast and Analysis (IDC #207849, July 2007)
- 2007 U.S. Mobile Imaging Survey (IDC #207847, August 2007)
- Worldwide Search and Discovery Software 2007–2011 Forecast (IDC #206148, March 2007)
- 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)
- U.S. Wireless Music Service 2007–2011 Forecast and Analysis (IDC #207304, June 2007)
- Worldwide and U.S. Portable Media Player 2007–2011 Forecast and Analysis (IDC #206016, March 2007)
- U.S. Mobile Music 2007–2011 Forecast (IDC #207275, June 2007)
- U.S. Wireless Music Service 2007–2011 Forecast and Analysis (IDC #207304, June 2007)
- Worldwide Video-Enabled PMP 2007–2011 Forecast and Analysis: Video to Go (IDC #208459, September 2007)
- Worldwide Archive and Hierarchical Storage Management Software 2007–2011 Forecast: Retention, Preservation, Optimization, and Reuse (IDC #206226, April 2007)

- Worldwide IT Security Software, Hardware, and Services 2007–2011 Forecast: The Big Picture (IDC #210018, December 2007)
- Worldwide Email Usage 2007–2011 Forecast: Resurgence of Spam Takes Its Toll (IDC #206038, March 2007)
- Worldwide Email Archiving Applications 2007–2011 Forecast and 2006 Vendor Shares: Storage Optimization, Mailbox Management, and Records Retention for eDiscovery and Compliance Drive Investments (IDC #206729, May 2007)
- Worldwide Compliance Infrastructure 2007–2011 Forecast: Compliant Information Infrastructure, Data Privacy, and IT Risk and Compliance Management Underpin Spending (IDC #209257, November 2007)
- Worldwide Enterprise Instant Messaging Applications and Management Products 2007–2011 Forecast: UC Mania Puts Spotlight on EIM (IDC #209596, December 2007)
- Worldwide Videogame Console Hardware and Software 2007–2011 Forecast and Analysis: Ready to Play a New Way (IDC #205659, February 2007)
- Worldwide Multifunction Peripheral 2007–2011 Forecast and Analysis (IDC #208293, September 2007)
- Worldwide IP PBX and Desktop Hardware IP Phone 2007–2011 Forecast: The Year of Unified Communications? (IDC #206112, March 2007)
- U.S. Residential VoIP Services 2007–2011 Forecast: The Race Is Just Beginning (IDC #208334, September 2007)
- U.S. Residential VoIP Handset 2006–2010 Forecast and Analysis (IDC #204690, December 2006)
- Demystifying the Digital Oilfield (IDC #EI202344, July 2006)
- Worldwide Server Power and Cooling Expense 2006–2010 Forecast (IDC #203598, September 2006)
- Moving Beyond the Hype: The Future of the Intelligent Grid (IDC #EI202543, July 2006)
- Intelligent Utilities: The Future of Electric Grids (IDC #EIOS01P, November 2007)
- Worldwide Disk Storage Systems 2007–2011 Forecast Update (IDC #209490, December 2007)
- Worldwide Hard Disk Drive 2007–2011 Forecast Update (IDC #209583, November 2007)
- Worldwide NAND Flash Demand and Supply 2Q07–4Q08 and 2007–2011 Update (IDC #208784, October 2007)
- Worldwide DRAM Demand and Supply 2Q07–4Q08 and 2007–2011 Update (IDC #208785, October 2007)

ADDITIONAL DATA SOURCES

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

ENDNOTES

[i] www.emc.com/digital_universe.

[ii] Avogadro's number, in atoms, refers to the number of atoms in a mass the size of the substance's mass in grams. For more information, see http://en.wikipedia.org/wiki/Avogadro_constant.

[iii] An exabyte is a billion gigabytes, while a gigabyte is a billion bytes. A byte is composed of 8 digital bits, each either a zero or a one. A byte typically encodes one letter, number, or special character in the Western alphabet or number system.

[iv] http://en.wikipedia.org/wiki/Dark_matter.

[v] Charles Babcock, "Data, Data, Everywhere," *InformationWeek*, January 9, 2006.

[vi] Graham P. Collins and author interviews with CERN staff, "Large Hadron Collider: The Discovery Machine," *Scientific American*, February 2008.

[vii] The Personal Digital Footprint Calculator allows individuals to fill out a simple questionnaire to determine their own digital footprint. It can be seen at and downloaded from www.emc.com/digital_universe.

[viii] http://www.newstatesman.com/200610020022.

[ix] http://blog.wired.com/sterling/2007/12/burning-british.html.

[x] Maria Aspan, "How Sticky Is Membership on Facebook?" *The New York Times*, February 11, 2008.

[xi] http://www.monthlyreview.org/mrzine/brenner070907.html.

[xii] Hiawatha Bray, "Google Faces Order to Give Up Records," *The Boston Globe*, March 15, 2006.

[xiii] "Beijing Clinic Ministers to Online Addicts," MSNBC, July 2005, from Associated Press.

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 1,000 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

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