

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



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. Experts say more than a billion songs a day are shared over the Internet in MP3 format. Digital bits. London's 200 traffic surveillance cameras send 64 trillion bits a day to the command data center. Chevron's CIO says his company accumulates data at the rate of 2 terabytes – 17,592,000,000,000 bits – a day. 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 $1,288 \times 10^{18}$ 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.

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.



To read the full report, go to www.emc.com/about/destination/digital_universe.

COPYRIGHT NOTICE

External Publication of IDC Information and Data. Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from IDC. A draft of the proposed document should accompany any such request. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/offices.

Copyright 2007 IDC. Reproduction is forbidden unless authorized. All rights reserved.



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