



- Technical Library
- Customer Training
- Customer Service
- Partnerships & Alliances
- Contact EMC
- Site Map

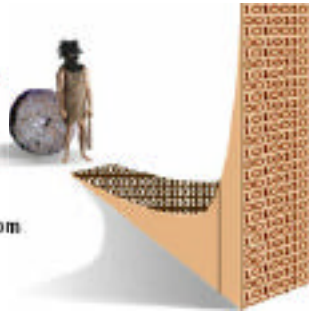
Advanced Search

In-Depth Coverage

Information Generation

Berkeley study measures gargantuan information boom

By Eric Woodman
Woodman_Eric@emc.com



October 19, 2000 — It's taken 300,000 years for humans to accumulate 12 exabytes of information. It will take just 2.5 more years to create the next 12 exabytes, according to a new study produced by a team of faculty and students at the School of Information Management and Systems at the University of California at Berkeley.

The study, sponsored by EMC and titled "How Much Information?" is the most comprehensive analysis of the world's digital and non-digital data volume to date. To reach their conclusions, the researchers looked at a range of media and recording methods — including print, broadcast, film, letters and the Internet - and focused on the year 1999 to determine the rate of information generation. The study for the first time makes it possible to compare growth trends for different media using one universal standard — the terabyte.

What's an Exabyte?
 It's 1 000 000 000 000 000 000 bytes, or 50,000 times the volume of the Library of Congress.
Source: The Internet Archive

How much is 12 Exabytes?
 Stored on floppy disks, the information would make 24 million miles high.
Source: USA Today

"What we are seeing is a phenomenal boom in the production of information," said Hal Varian, dean of the school and co-author of the study. "One interesting finding is the degree to which individuals, rather than organizations, are responsible for generating data. We not only have mass production of information, but also production of information by the masses."

As this individual-based data is created, business opportunities are expected to emerge. For example, the study found that a full 55 percent of the world's digital information is confined to single-user personal computers, compared with the 16 percent that is stored in corporate data warehouses. Privacy concerns aside, Varian believes many will look to specialized data storage businesses as the 250 megabytes of data generated on average by each person on Earth doubles every year.

"It comes down to who can do it best," he said. "Would you rather keep your entire collection of family photos on a PC hard drive or on a secure site managed by a professional? One thing is certain, the demand for information management will be tremendous."

The Berkeley researchers noted several key trends:

- There were 1.5 exabytes of information created in 1999. Varian projects that number will double every year for the foreseeable future.
- Digital information is driving the boom. Not only is digital information the largest type of data

For much more on the "How Much Information?" study, visit the study's [website](#).

- Read coverage of the study from
- [USA Today](#)
 - [The Economist](#)

View EMC's [news release](#)

View Berkeley's [news release](#)

FROM THE STUDY

Some "[sound bytes](#)" from the University of California, Berkeley study:

- **The Democratization of Data.**
 Individuals produce significant amounts of non-digital information. As photos and videos move to digital formats, households will have to manage terabytes of data.
- **Magnetic Migration.**
 Print and film content is rapidly moving to magnetic and optical storage. This is true for professional use now, and will become increasingly true at the level of individual users.
- **Terror of Terabytes.**
 One terabyte, the smallest practical measure used for the project, is a million megabytes, which is equivalent to the textual content of a million books. An exabyte, used to report the final results, is a billion gigabytes.
- **Dominance of Digital.**
 Ninety-three percent of the information produced each year is stored in digital form. Hard drives in stand-alone PCs account for 55 percent of total storage shipped each year.
- **Immensity of Images.**
 More than 80 billion photographs are taken every year, which would require more than 400 petabytes to store -- representing more than 80 million times the storage requirements for text.
- **Convenience of copies.**
 There is a lot of redundancy both across and within media. A newspaper, for example, is created using digital technology, printed on paper, then archived on microfilm. Estimates of "unique" information can only be taken as approximate.

Would you like to be notified by e-mail when EMC issues a new news release? [Subscribe to EMC E-News.](#)

produced, but traditionally non-digital items — such as books, music, films and medical records — are being digitized as a rapid pace.

- While the Internet is growing rapidly, "stock" web pages account for just 21 terabytes of storage. Far more significant is the impact of e-mail, which contributes more than 500 times more data to the total pool each year than the volume generated by new web pages.
- The total time people spend accessing data has remained nearly constant throughout the 1990s. What's changed is that people are spending more time on interactive pastimes — like computer games, videos and the Internet — at the expense of time spent reading print items.

Varian speculates that significant advances will have to be made in information management before mankind can fully benefit from the information explosion. "Will we drown in a sea of information, or can we develop tools to help us swim?" he asked.

"The difficulty will be in managing this information effectively," said Varian. "This is no easy task. Our ability to store and communicate information has far outpaced our ability to search, retrieve and present it. Information management may turn out to be one of the major challenges of the new century."

