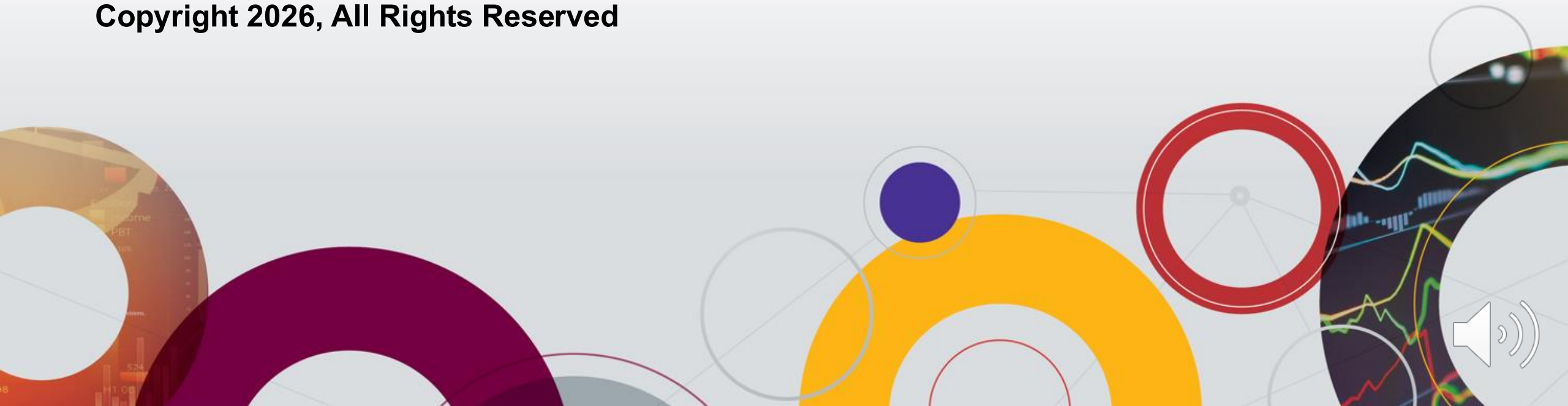


2.2 Data Storage in Computers

10x10 – Introduction to Biomedical Informatics & Artificial Intelligence
William Hersh, MD, FACMI, FAMIA, FIAHSI
Copyright 2026, All Rights Reserved



Digital concepts – overview

Virtually all modern computers are digital, as opposed to analog

Digital means that

- Most fundamental unit is discrete
- Unit is the binary digit or bit, which can assume the values 0 or 1
 - Or off/on, false/true, etc.
- A byte is a sequence of 8 bits (and can take on 2^8 or 256 values)
 - Multiple bytes represented by prefixes kilo- (1000 bytes), mega- (1000 kilobytes), etc.

Digital concepts

In contrast, analog systems have values on continuous scale

Examples of digital vs. analog

- Timekeeping
- Calculators vs. slide rules
- Audio – from vinyl to CDs and other digital forms



Digital concepts – data representation

Bit sequences are used to represent numbers, text, program instructions, images, video, genomes, etc.

All these sequences are stored in memory and structure determines what the data represents

Data representation – numbers

Integer numbers are represented by straight sequences of bits

- Eight bits can represent 256 numbers, such as 0 to 255 or -128 to +127
- Larger integers require more bits

Floating point numbers can be represented in variable ways; one approach is to have two sequences for digits on either side of decimal point

- e.g., 5.2

Important for system to interpret consistently

Data representation – text

Text is represented by codes

Most common is ASCII, which represents alphanumeric characters in 7 bits (128 unique characters)

Most PC systems extend ASCII with an 8th bit

- How many unique characters does this allow?
- Is that enough?

Text representation (cont.)

But even 256 characters is not enough for Chinese, Japanese, Korean, Hebrew, Arabic, Greek, Russian, etc.

Unicode allows representation of larger character sets

- Uses 8, 16, and 32-bit representations
- Designated by UTF-8, UTF-16, and UTF-32 respectively
- Is this sufficient for all human languages?

Even allows their use for emojis 🤔

Data representation – images

Images are represented by pixels (picture elements)

- Each pixel has one (B&W) or more (color) bits at each level of depth
 - 8-bit color has 256 possible colors (or gray-scale)
- How much memory does a 24-bit color, 3000 x 2000 pixel image use?
 - $3000 * 2000 = 6,000,000$ pixels (or 6 megapixels [MP], a common unit of measurement in digital photography)
 - $6,000,000 \text{ pixels} * 24 \text{ bits/pixel} = 144,000,000 \text{ bits} = 18,000,000 \text{ bytes} = 18 \text{ megabytes (MB; usually compressed in standard image, e.g., JPEG)}$

Is a picture worth a thousand words? Compare storage sizes

Text

- For text, each character is one byte
- A single page may have an average of 50 characters per line over its 60 lines, taking up about 3,000 bytes

Image

- A high-quality display of a page requires about 1200 by 800 pixels, taking up (in B&W) around 120,000 bytes (before compression)
- Adding 8-bit color would increase the size to nearly one million bytes (before compression)

Can achieve reduction in size of images with compression

Types of compression

- Lossless – no resolution lost upon reconstruction
 - Best achievable compression around 2:1-3:1
- Lossy – eye less sensitive to certain changes (e.g., hue or color), so can lose some depth of hue with little or no notice
 - In medical images, up to as high as 20-30 to 1 for x-rays, and 6 to 1 for CT scans

Classic and newer image compression methods

- JPEG – world-wide standard, lossy
- Older lossless – TIFF and PNG
- Newer lossless and lossy – HEIC and WebP (Alvi, 2023)

Digital conversions

1000 bytes = 1 kilobyte (or 1K)

1000 kilobytes = 1 megabyte (or 1 meg)

1000 megabytes = 1 gigabyte (or 1 gig)

1000 gigabytes = 1 terabyte

1000 terabytes = 1 petabyte

1000 petabytes = 1 exabyte

1000 exabytes = 1 zettabyte

1000 zettabytes = 1 yottabyte = 10^{24} bytes

In early days of computing, “binary” conversion was used such that 1024 byte = 1 K, 1024 K = 1 M, etc.

- This system now called kibibyte, mebibyte, etc.



Sizes of storage media

Floppy disks held 1440 kilobytes (1440K)

CD disks hold 640-700 megabytes

DVD disks hold 4.7 GB (current standard) and up to 50 GB (Blu-Ray)

Hard disks ever increasing in size, up to many terabytes

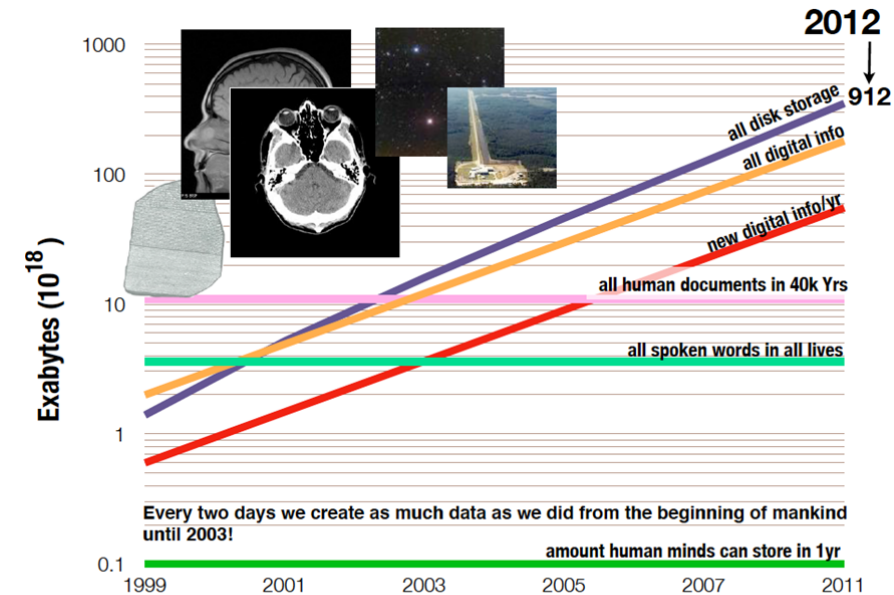
- Originally magnetic
- Growing use of flash or solid-state (used in USB or thumb drives)

Logarithmic growth of data

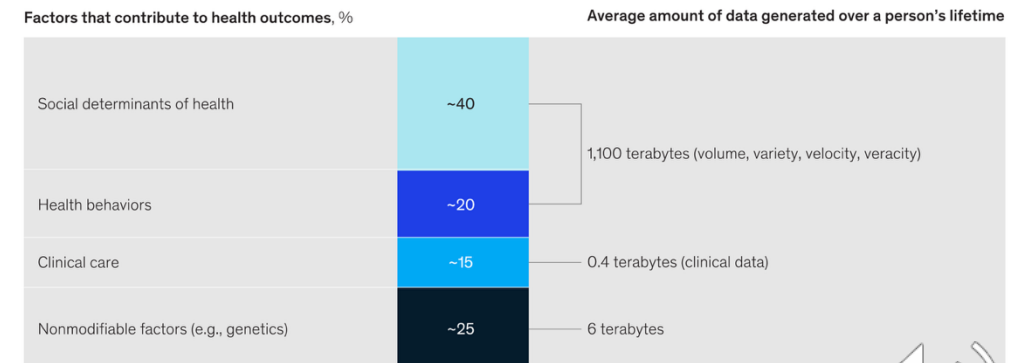
For digital but not analog or human data (Lesk and others)

Average person likely to generate 1.1 PB of data over lifetime, only a tiny fraction from healthcare (Singhal, 2018)

Healthcare organizations saw explosive data growth rate of 878% between 2016-2019 to average of 8.41 petabytes (Donovan, 2019)



Sources: Lesk, Berkeley SIMS, Landauer, EMC, TechCrunch, Smart Planet



“Global datasphere” is growing

Data worldwide continues to grow (Petrosyan, 2025)

A day in data (Cassidy, 2019) – by 2025, 463 EB generated each day

