

UC Santa Barbara

Core Curriculum-Geographic Information Systems (1990)

Title

Unit 03 - Introduction to Computers

Permalink

<https://escholarship.org/uc/item/9n10v5m2>

Authors

Unit 03, CC in GIS

National Center for Geographic Information and Analysis

Publication Date

1990

Peer reviewed

UNIT 3 - INTRODUCTION TO COMPUTERS

UNIT 3 - INTRODUCTION TO COMPUTERS

- [A. INTRODUCTION](#)
- [B. COMPUTER DATA](#)
 - [Binary notation](#)
 - [Bits and bytes](#)
 - [ASCII coding system](#)
- [C. COMPUTER HARDWARE](#)
 - [Central processing unit \(CPU\)](#)
 - [Memory](#)
 - [Peripherals](#)
- [D. DATA STORAGE](#)
 - [Storage media](#)
 - [Fixed disks](#)
 - [Dismountable devices](#)
 - [Volumes](#)
 - [Files](#)
- [E. SOFTWARE](#)
 - [Programs](#)
 - [Operating systems](#)
 - [Compilers and languages](#)
 - [Applications programs](#)
- [F. EDITORS AND WORD PROCESSORS](#)
- [G. DATABASES](#)
 - [Functions of a database](#)
 - [Three types of database](#)
- [H. SPREADSHEETS](#)
- [I. STATISTICAL PACKAGES](#)
- [REFERENCES](#)

- [EXAM AND DISCUSSION QUESTIONS](#)

This unit provides a brief introduction to computer hardware and software. We have included this unit to help those who are teaching students with no computer background. However, any introductory course in the use of micro-computers is likely to have covered this material already. Binary notation is introduced here. A knowledge of the binary numbering system and conversion to decimal is needed only for Units 35, 36 and 37 but it is useful for students to be aware of this fundamental topic.

UNIT 3 - INTRODUCTION TO COMPUTERS

[A. INTRODUCTION](#)

- the environment in which a GIS operates is defined by:
 - hardware - the machinery, including:
 - a host computer
 - ranging from a stand-alone microcomputer to a large mainframe supporting many users
 - several devices for handling input and output
 - software
 - the programs that tell the computer what to do
 - the data the programs will use
- this unit provides a brief overview of computer hardware and software so that students will have a basic understanding of how computers operate and will recognize some of the common computer terminology
 - important topics are covered in greater detail in later units

[B. COMPUTER DATA](#)

- computer data is coded, manipulated and stored by use of an exclusive two-state condition
 - in English such two-state forms of information can include yes/no, on/off, open/closed, hole/no hole
 - in simple electronic terms this two-state condition can be translated for the computer into "switch open/switch closed", meaning that "there is electricity passing through the circuit/there is no electricity passing through the circuit"
 - note that one of the two exclusive states always exists
- if one switch provides two different datum, how much data can we obtain from two switches?
 - four - there are four combinations of open and closed switches

[Binary notation](#)

- in computer terminology, this two state condition is represented in binary notation by the use of 1s and 0s
- thus, two switches produce four codes - 00, 01, 10, 11
 - three switches produce eight codes - 000, 001, 010, 011, 100, 101, 110, 111
- in mathematical terms:
 - 1 binary digit provides $2^1 = 2$ alternatives
 - 2 binary digits provide $2^2 = 4$ alternatives
 - 3 binary digits provide $2^3 = 8$ alternatives
 - 8 binary digits provide $2^8 = 256$ alternatives

Bits and bytes

- each binary digit is called a bit
 - the complexity of computer circuitry is described in terms of the number of bits that can be transmitted simultaneously
 - this is determined by the number of wires that run parallel to one another on the circuit-boards
 - current PCs use 8, 16 and 32 bit paths
- a group of 8 bits is called a byte
 - bytes are the standard unit of measurement of computer data

ASCII coding system

- to maximize efficiency, most computers store data in their own internal formats
 - however, transfer of data requires the use of standard codes which are understood by all systems
- the most successful standard is ASCII (pronounced ass- key)
 - ASCII originated well before computer communication as a code for Teletypes
 - ASCII assigns the numbers 0 through 127 to 128 characters, including the upper and lower case alphabets, numerals 0 through 9 and various special characters
 - 128 different patterns can be generated using 7 bits in different combinations of on and off
 - any ASCII character can therefore be coded with 7 bits
 - in practice, 8 bits (one byte) are used, the extra bit may be used to extend the code to 128 extra characters, or it simply may be redundant
- by using binary notation, these codes can be converted into decimal numbers
 - counting from the right, the 8 bits are numbered 0 through 7, and signify as follows:

Bit: 7 6 5 4 3 2 1 0 128s 64s 32s 16s 8s 4s 2s units

- e.g. the combination 01010101 is no 128s, one 64, no 32s, one 16, no 8s, one 4, no 2s and one unit i.e. $64+16+4+1 = 85$

- in the ASCII code system, code number 85 is an upper case U
 - thus to store a U, the system stores a byte with the bit pattern 01010101
- in ASCII, characters 0 through 32 often perform special functions
 - e.g. character 7, 00000111, is the BEL character and rings a bell if received by many terminals or devices
 - e.g. character 12, 00001100, is the FF character and produces a form feed (new page) if received by many printers
- computer files which contain information coded in ASCII are easily transferred and processed by different computers and programs
 - such files are often called "ASCII" or "text" or "coded" files
 - ASCII characters are the dominant basis for communication between different systems, and communication with peripherals
- files which are not ASCII are often coded in "binary" and generally can be processed or understood only by specific programs

C. COMPUTER HARDWARE

- computers consist of several different hardware components

Central processing unit (CPU)

- the central processing unit is the essential component of a computer because it is the part that executes the programs and controls the operation of all the hardware
 - powerful computers may have several processors handling different tasks, although there will need to be one central processing unit controlling the flow of instructions and data through the subsidiary processors
- the CPUs of PCs are based on a series of processors or "chips" from Intel
 - "PC" models use the 8088 (8 bit)
 - "AT" models use the 80286 (8/16 bits)
 - current high powered machines use the 80386 (full 16 bits) and 80486
- the Macintosh CPUs are based on the 68000 series of chips from Motorola

Memory

- memory stores input for and output from the CPU as well as the instructions that are followed by the CPU
- the amount stored is measured in bits, bytes, Kbytes (K, Kb, 103 bytes), Megabytes (Mb, 106 bytes), Gigabytes (Gb, 109), Terabytes (Tb, 1012)
- there are two kinds of memory:
- main memory (or internal or primary memory) is essential for the operation of the computer, all data and

instructions must be in main memory first before it can be processed by the computer

- most costly memory
 - in the form of microchips integrated with the computer's central processor
 - fastest access - any byte can be accessed equally rapidly (random access, hence it is called RAM)
 - temporary - since data and instructions are stored in main memory as electrical voltages, power failures cause the loss of all data in main memory
 - ranges from several hundred Kbytes for typical PC to many Megabytes for mainframes
- secondary memory (or auxiliary memory or secondary storage) is used for large, permanent or semi-permanent files
 - GIS programs and data generally require very large amounts of storage
 - data storage is covered after this overview of the components of computers

Peripherals

- peripherals refer to all the other devices attached to computers that handle input and output
 - input devices include keyboards, mice, trackballs, digitizers, disk drives
 - output devices include screens, printers, plotters
 - those devices important to GIS are examined in later units

D. DATA STORAGE

Storage media

- computers can use several different media for storing information
 - needed to store both raw data and programs
- media differ by
 - storage capacity
 - speed of access
 - permanency of storage
 - mode of access
 - cost

Fixed disks

- most costly memory next to main/internal memory is fixed disk memory
- ranges from 10 Megabytes for typical PC to hundreds of Gigabytes in large "disk farms"
- random access but slower than internal memory
- permanent (i.e. does not disappear when power is turned off), though data can be erased and modified

Dismountable devices

- dismountable devices can be removed for storage or shipping, include:
 - floppy diskettes
 - up to 1.44 Megabytes for PC - random access
 - magnetic tapes
 - tens of Megabytes for standard tape
 - access is sequential, not random
 - can take minutes to reach a particular set of data on the tape, depending on where it is stored
 - optical compact disks (CDs)
 - around 250 Megabytes per CD
 - random access, but the delay in reaching a given item of data may be 1 second or more

Volumes

- a volume is a single tape, CD, diskette or fixed disk, i.e. a physical unit of storage

Files

- a file is a logical collection of data - a table, document, program, map
- many files can be stored on a single volume
- files are given names
 - the rules for naming files vary among types of systems
- the computer operating system keeps track of files stored in a volume by using a table called a directory
 - files are identified in the directory by name, size, date of creation and often type of contents
- files are often organized into subdirectories so that the user can group files under specific topics

E. SOFTWARE

Programs

- a program is a sequence of related instructions, performed one step at a time by the CPU to accomplish some task
 - programs determine how computers respond to input, what will be displayed and output
- there are three types of programs: operating systems, language interpreters and compilers and applications programs

Operating systems

- an operating system (OS) is the software which controls the operation of the computer from the moment it is turned on or "booted"
 - the OS controls all input and output to and from the peripherals as well as the operation of other programs
 - allows the user to work with and manage files without knowing specifically how the data is stored and retrieved
- in multi-user systems, operating systems manage user access to the processor and peripherals and schedule jobs
- common operating systems include:
 - IBM PCs and clones use MS-DOS (often called DOS), although there is some movement to OS/2
 - UNIX (and similar operating systems such as AIX, XENIX) is the dominant operating system for workstations
 - mainframes commonly use proprietary operating systems developed by their manufacturers - VMS on DEC's VAX series, PRIMOS on Prime, CMS on IBM mainframes, etc.
- although functions performed by operating systems are similar, it can be very difficult to move files or software from one to another
 - many software packages run under only one operating system, or have substantially different versions for different operating systems

Compilers and languages

- since computers operate on electricity and binary operations, all instructions executed by computers must be provided to the CPU in machine code
 - however, humans do not have to interact with computers at this level
- programs can be written in very specialized languages, called assemblers, which allow programmers to take advantage of the specific capabilities of particular machines by addressing the basic operations directly
 - these languages are very cryptic and very difficult to use
 - they are also system specific and cannot be transported from one type of computer to another
- most programs are created using standard high level languages such as C, Pascal, FORTRAN, BASIC which are common across most computer systems, from micro to mainframe
 - such programs are referred to as source code
 - these languages generally use English words and familiar mathematical structure
- a compiler is a program designed to convert a program written in a high level language to the machine instructions of a specific computing system or "platform"
 - the output of a C compiler for the IBM PC has almost nothing in common with the output of a C compiler for a VAX mainframe

although high level languages are generally used in the development of application packages such as GIS, it is normally compiled for specific platforms before distribution to the public

- this is done to protect the commercial interests of the developer

Applications programs

- applications programs are programs used for all purposes other than performing operating system chores or writing other programs
 - includes GIS, word processors, spreadsheets, statistics packages and graphics programs, airline reservation systems, payroll systems

F. EDITORS AND WORD PROCESSORS

- are packages designed to modify or edit the contents of files
- are most often used to edit written text or programs
 - editing and creation of files of numerical data is best done with the special purpose editors found in database packages or spreadsheets (see sections G and H)
- editors and word processors are increasingly WYSIWYG ("what you see is what you get")
 - the screen shows a picture of the contents of the file at all times
 - well-known word processors for the IBM PC include Wordstar, WordPerfect and Microsoft Word
- linkage to a printer is essential so that the user can obtain "hard copy" of a file's contents
- many mainframes offer their users several editors
 - unfortunately there is little standardization of editors
- an editor is the most important system to learn after the operating system
 - it is difficult to make much effective use of a system without one

G. DATABASES

- are packages designed to create, edit, manipulate and analyze data
- to be suitable for a database, the data must consist of records which provide information on individual cases, people, places, features, etc.
- each record may contain several fields each of which contains one item of information
 - the number and interpretation of the fields must be constant for each class of records
 - e.g. each record in the class of "streets" may contain fields for name, length, surface, type.
 - field contents can be of many types - numeric or text, fixed or variable length

- there can be several classes of records in a database
 - e.g. an airline reservation database might have the following classes of records and associated items: passengers: name, phone, flight numbers aircraft: type, registration number, number of seats crew: names of pilot, copilot, cabin crew, home city flight: number, departure and arrival times, aircraft

Functions of a database

- creating and editing records, using customized screens
- printing reports (summarizes of groups of records), using customized report forms, including subtotals and totals
- selecting records based on user-specified rules
- updating records based on new information
- linking records, e.g. to determine arrival time for a passenger by linking the passenger's record with the correct flight record

Three types of database

- network, hierarchical and relational are different ways of modeling data within a database
- although all three are used, the relational model has been most successful within GIS
 - it is discussed at length later in the course
 - well-known relational database management systems (RDBMSs) include dBase, Oracle, Info
 - many of these have been used in specific GISs
- many databases use the same language, SQL (Standard Query Language), for formulating queries

H. SPREADSHEETS

- are systems which allow the user to work with numerical data in tabular form
- column and row totals, percentages etc. are automatically updated as data items are changed
- Lotus 1-2-3 is a well-known spreadsheet for the IBM PC

I. STATISTICAL PACKAGES

- offer a range of types of statistical analysis
- data is primarily numerical

may include:

- database functions, such as editing, printing reports
- capabilities for graphic output, particularly graphs but many also produce maps
- common mainframe packages are SAS, SPSS, BMD
 - available over a wide range of operating systems
 - some have been "ported" to (rewritten for) the IBM PC
- numerous other packages have been developed specifically for the PC DOS environment
- S is a commonly available statistical package for UNIX

REFERENCES

Maguire, D.J., 1989. Computers in Geography, John Wiley and Sons, Inc., New York.

Current reviews and comparisons of different hardware and software are published frequently, particularly for the DOS environment in magazines such as Byte and PC Magazine.

Numerous texts are available at various levels of sophistication for operating systems, editors, compilers and common applications programs.

EXAM AND DISCUSSION QUESTIONS

1. Compare the data storage needs of (a) the data which will be transmitted by the EOS satellites of the 1990s, which will generate approximately 1 Terabyte/day, (b) the US Bureau of the Census's TIGER files of street networks, which amount to about 10 Gigabytes and are updated every 10 years, and (c) a database of 100 Megabytes created for use in a one-time environmental impact study
2. "User expectations about data volumes rise at least as rapidly as the capacity of available storage devices". Discuss.
3. Why do you think the computer industry has been unable to agree on a common operating system? or single source language?
4. Describe the functional differences between databases, spreadsheets and statistical packages. Which would be more useful for (a) research in a university department, (b) administrative record-keeping in a small business, (c) personal budget planning?

Last Updated: August 30, 1997.