

CAPITOLO 4: Formati dei dati

The Architecture of Computer Hardware, Systems Software & Networking: An Information Technology Approach

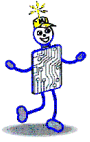
5th Edition, Irv Englander

John Wiley and Sons ©2013

Diapositive realizzate da Angela Clark, University of South Alabama

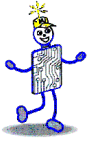
Diapositive per la 4° edizione realizzate da Wilson Wong, Bentley University

Diapositive per il CLEI tradotte e adattate da Gianluca Amato, Univ. CH-PI



Formati dei dati

- Computer e dispositivi simili
 - Elaborano e memorizzano tutti i tipi di dati in formato binario
- Comunicazione umana
 - Include linguaggio, immagini e suoni
- Formati dei dati
 - Specifiche per convertire i dati in forma usabile dal computer
 - Definisce i vari modi in cui i dati umani possono essere rappresentati, memorizzati ed elaborati da un computer



Sorgenti dei dati

- Sorgenti digitali
 - I dati sono entità discrete
 - Esempio: input da tastiera come A 1+2=3 math
 - La tastiera genera un codice numerico binario per ogni tasto
- Sorgenti analogiche
 - Dati continui come suoni o immagini
 - Richiede un dispositivo hardware per convertire i dati in numeri binari

A 1+2=3 math →

Input
device



Computer

1101000101010101...



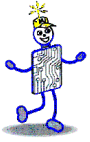
Tipici formati per dati

| Type of Data | Standard(s) |
|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Alfanumerico | Unicode, ASCII, EDCDIC |
| Immagine (bitmap) | <ul style="list-style-type: none">▪ GIF (graphical image format)▪ TIFF (tagged image file format)▪ PNG (portable network graphics) |
| Immagine (oggetto) | PostScript, JPEG, SWF (Adobe Flash), SVG |
| Suono | WAV, AVI, MP3, MIDI, WMA |
| Descrizione di pagina | Postscript PDF (Portable Document Format), HTML, XML |
| Video | Quicktime, MPEG-2, MPEG-4, WMV |



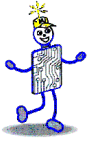
Rappresentazione interna dei dati

- Dipende da:
 - Complessità del tipo di dato
 - Tipo di elaborazione da eseguire
- Trade-off
 - Accuratezza e risoluzione
 - Semplice foto vs figura in un libro di illustrazioni
 - Compattezza (memorizzazione e trasmissione)
 - Maggiori quantità di dati consentono di avere maggiore accuratezza e precisione
 - La compressione consente di rappresentare i dati in forma più compatta
 - Metadati: dati che descrivono o interpretano il significato dei dati



Rappresentazione interna dei dati

- Facilità di manipolazione
 - Elaborazione di semplici clip audio vs elaborazione di audio ad alta-fedeltà
- Standardizzazione
 - Formati proprietari per memorizzare ed elaborare dati (WordPerfect vs. Word)
 - Standard di fatto: standard proprietari basati sulla ampia diffusione (PostScript)



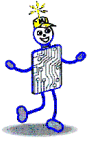
Tipi di dati: Numerici

- Usati per i calcoli matematici
 - Somma, sottrazione, moltiplicazione, divisione
- Tipi
 - Interi
 - Reali (contengono il punto decimale)
- Non vengono trattati in queste slide



Tipi di dati: Alfanumerici

- Alfanumerico:
 - Caratteri: b T
 - Cifre: 7 9
 - Segni di punteggiatura: ! ;
 - Caratteri ad uso speciale: \$ &
- Caratteri numerici vs numeri
 - Entrambi vengono inseriti come caratteri
 - Il computer converte i caratteri in numeri per l'elaborazione
 - Esempio: metodo Integer.parseInt in Java
 - Trattati come caratteri se elaborati come testi
 - Esempi: numero di telefono, CAP



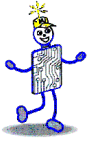
Codici alfanumerici

- Scelta arbitraria di bit per rappresentare i caratteri
 - Consistenza: dispositivi di input e output devono usare lo stesso codice
 - Il valore dei numeri binari usati per rappresentare i caratteri corrisponde all'ordine dei caratteri nell'alfabeto
 - Facilita ordinamento e ricerca



Rappresentare i caratteri

- ASCII: sistema di codifica più comune
- EBCDIC: usato da mainframe IBM (ormai del tutto in disuso)
- Unicode: sviluppato per essere utilizzato in tutto il mondo



ASCII

- Sviluppato dall'ANSI (American National Standards Institute)
- Rappresenta
 - Alfabeto latino, cifre numeriche arabe, simboli di punteggiatura standard
 - Più un piccolo insieme di accenti e altri caratteri speciali europei
- ASCII
 - Codice a 7-bit: 128 caratteri

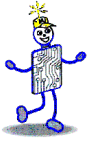


ASCII Reference Table

| MSD LSD | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------------|-----|-----|----|---|---|---|---|----------|
| 0 | NUL | DLE | SP | 0 | @ | P | | p |
| 1 | SOH | DC1 | ! | 1 | A | Q | a | W |
| 2 | STX | DC2 | " | 2 | B | R | b | r |
| 3 | ETX | DC3 | # | 3 | C | S | c | s |
| 4 | EOT | DC4 | \$ | 4 | D | T | d | t |
| 5 | ENQ | NAK | % | 5 | E | U | e | u |
| 6 | ACJ | SYN | & | 6 | F | V | f | v |
| 7 | BEL | ETB | ' | 7 | G | W | g | w |
| 8 | BS | CAN | (| 8 | H | X | h | x |
| 9 | HT | EM |) | 9 | I | Y | i | y |
| A | LF | SUB | * | : | J | Z | j | z |
| B | VT | ESC | + | ; | K | [| k | { |
| C | FF | FS | , | < | L | \ | l | |
| D | CR | GS | - | = | M |] | m | } |
| E | SO | RS | . | > | N | ^ | n | ~ |
| F | SI | US | / | ? | O | _ | o | DEL |

74₁₆

111 0100



Due tipi di codici

- Caratteri stampabili
 - Producono un simbolo su schermo o stampante
- Caratteri di controllo
 - Controllano la posizione dell'output
 - VT: tabul. vert. ▫ LF: avanzamento riga
 - Causano il verificarsi di una azione
 - BEL: beep audio ▫ DEL: cancella carattere
 - Comunicano lo stato tra il computer la periferica
 - ESC: cambia il significato dei caratteri successivi, consentendo possibili estensioni del codice ASCII



EBCDIC

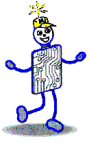
- Extended Binary Coded Decimal Interchange Code sviluppato dalla IBM
 - Usato essenzialmente da mainframe IBM o compatibili (ormai obsoleto)
 - Disponibile software di conversione a/da ASCII
 - Comune nei vecchi dati archiviati
 - Codice dei caratteri differenti dall'ASCII

| | ASCII | EBCDIC |
|-------|------------------|------------------|
| Space | 20 ₁₆ | 40 ₁₆ |
| A | 41 ₁₆ | C1 ₁₆ |
| b | 62 ₁₆ | 82 ₁₆ |



Unicode

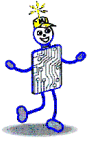
- Codice a 32 bit
- ASCII e Latin-1 sono sottoinsiemi di Unicode
 - Valori da a 0 a 255 del codice Unicode
- Multilingue: definisce codici per
 - Pressoché qualunque alfabeto basato su fonemi
 - Un nutrito insieme di caratteri ideografici per cinese, giapponese e coreano
 - Accentu e gruppi sillabici richiesti da alcune lingue
- Consente ad un programma di funzionare con qualunque lingua



Collating Sequence

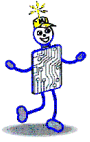
- Con “collating sequence” si intende l’ordinamento alfabetico dei caratteri
- In ASCII, numeri vengono prima nell’ordinamento; in EBCDIC, vengono dopo
- Ordinamento ASCII per una stringa di caratteri

| Letters | | | | | | | Numeric Characters | | | | | |
|---------|---|---|---|---|---|---|--------------------|-----|------|------|-----|------|
| Adam | A | d | a | m | | | 1 | 011 | 0001 | | | |
| Adamian | A | d | a | m | i | a | n | 12 | 011 | 0001 | 011 | 0010 |
| Adams | A | d | a | m | s | | | 2 | 011 | 0010 | | |



Dati visuali

- Video, fotografie, immagini biometriche, figure, icone, disegni, grafici
- Due approcci
 - Immagini *bitmap* o *raster* per foto o dipinti con variazioni continue di colore
 - Immagini a *oggetto* o *vettoriali* composte da *oggetti grafici* come linee e curve definiti in maniera geometrica
- Le differenze includono:
 - Qualità dell'immagine
 - Spazio richiesto per l'immagazzinamento
 - Tempo per la trasmissione
 - Facilità di modifica



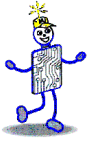
Immagini bitmap

- Usate per immagini realistiche con variazioni continue di colore, forma, ombra.
- Preferite quando l'immagine contiene una grande quantità di dettagli e non sono richieste elaborazioni complesse.
- Dispositivi di input:
 - Scanner
 - Fotocamera e videocamera digitale
 - Dispositivi di input come mouse o penne ottiche
- Gestite da software di disegno o di manipolazione video.



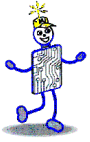
Tipi di immagini bitmap

- Monocromatiche: bianche o nere
 - 1 bit per pixel
- A scala di grigi: bianco, nero o 254 tonalità di grigio
 - 1 byte per pixel
- A colori: 16 colori, 256 colori, o 24-bit true color (16.7 milioni di colori)
 - 4, 8, e 24 bit rispettivamente



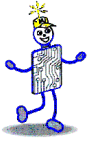
Memorizzare immagini bitmap

- Immagini spesso molto grosse
 - Esempio: 768 righe di 1024 pixel con 24 bit per ogni pixel ➡ ~2.4MB file
- La dimensione del file dipende da:
 - *Risoluzione* (numero di pixel)
 - Quantità di dettagli che influenza la chiarezza e la precisione dell'immagine
 - *Livelli*: numero di bit per ogni pixel
 - Palette: tecnica che si basa su un tabella di traduzione del colore che usa un codice (tipicamente a 8 bit) per ogni pixel invece della codifica effettiva (spesso a 24 bit per pixel)
 - Compressione dati

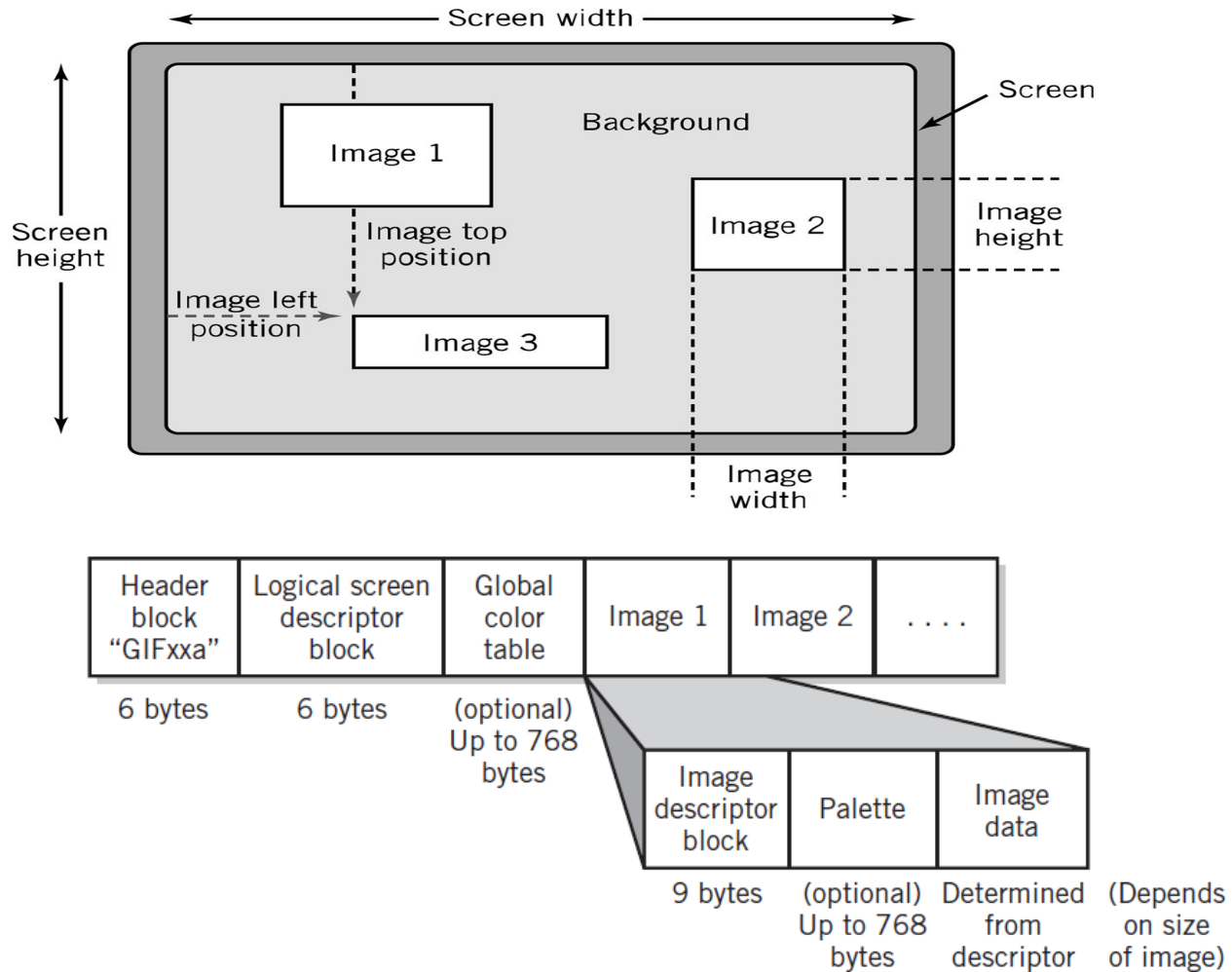


GIF (Graphics Interchange Format)

- Sviluppato inizialmente da CompuServe nel 1987
- GIF89a aggiunge le immagini animate
 - Consente di visualizzare delle immagine sequenzialmente a intervalli di tempo specificati
- Limitazione nel numero di colori: 256
- Immagini compresse con l' algoritmo LZW (Lempel-Zif-Welch)
- Preferite per disegni e figure con grossi blocchi di colore solido
- *Compressione lossless*



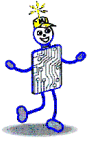
GIF (Graphics Interchange Format)





PNG (Portable Network Graphics)

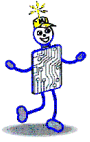
- Alternativa al formato GIF
- Può usare fino a 48 bit per pixel
- Può anche memorizzare la percentuale di trasparenza per ogni pixel, e fattori di correzione del colore per monitor e stampanti.
- Algoritmo di compressione più efficiente del GIF.



JPEG

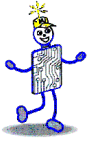
(Joint Photographers Expert Group)

- Consente più di 16 milioni di colori
- Utilizzato per immagini e disegni fotografici con molti dettagli
- Utilizza un algoritmo di *compressione lossy* che
 - Elimina dati per diminuire la dimensione del file e i tempi di trasmissione
 - Può ridurre la qualità dell'immagine, tendendo a sfocare i contorni netti



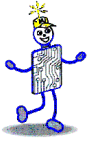
Immagini a oggetti

- Creati da software di *disegno* o prodotti come grafici dai fogli di calcolo
- Composti da linee e forme in vari colori
- Il computer interpreta la descrizione geometrica per produrre l'output
- Lo spazio di memoria occupato dipende dalla complessità dell'immagine
 - Numero di istruzioni per la creazione di linee, forme, riempimenti
- Film come Shrek e Toy Story usano immagini a oggetti.



Immagini a oggetti

- Basate su formule matematiche
 - È facile spostare, ridimensionare, ruotare gli oggetti separatamente
- Richiedono meno spazio di memorizzazione rispetto alle immagini bitmap
- Non posso rappresentare foto o dipinti
- Non possono essere visualizzate o stampate direttamente
 - Devono essere prima convertite in immagini bitmap perché i dispositivi di output (tranne i plotter a penna) sono dispositivi bitmap



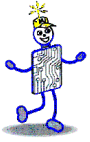
PostScript

- *Linguaggio di descrizione di pagina*: elenco di comandi e procedure che descrivono gli oggetti che devono essere stampati in una pagina
 - Memorizzati come testo ASCII o Unicode
 - Un programma interprete nel computer o nella stampante legge il PostScript e genera una immagine bitmap
- Supporto per font scalabili
 - I font sono descritti per via matematica come gli altri oggetti



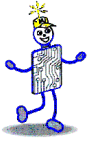
Immagini bitmap vs vettoriali

| Bitmap (Raster) | Oggetti (Vettori) |
|----------------------------------------------------------------|------------------------------------------------------------------|
| Matrice di pixel | Forme definite geometricamente |
| Qualità fotografica | Disegni complessi |
| Programma di pittura digitale | Programma di disegno |
| Notevoli richieste di spazio di immagazzinamento | Notevoli richieste computazionali |
| Ingrandire l'immagine produce bordi seghettati | Gli oggetti si possono ingrandire senza perdere qualità |
| Risoluzione di output limitata dalla risoluzione dell'immagine | Risoluzione dell'output limitata solo dal dispositivo utilizzato |



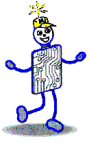
Video

- Richiedono una enorme quantità di spazio
 - Una videocamera che produce immagini FullHD 1920x1080 pixel true-color a 30 frame/sec ➡ 180 MB di dati/sec
 - Clip da 1 minuto ➡ 10.4 GB di spazio
- Opzioni per ridurre la dimensione dei file: diminuire la risoluzione dell'immagine, limitare il numero di colori, ridurre il numero di frame (immagini) al secondo.
- Il formato video è determinato da un codec, abbreviazione di encoder/decoder.



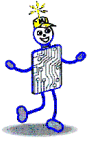
Video

- Codec standard più utilizzati: MPEG-2, MPEG-4 e H.264
 - I dati possono essere compressi fino ad arrivare a soli 10-60 MB al minuto.
- Un formato “container” serve come sovrastruttura per contenere il video vero e proprio ma anche tracce audio alternative, sottotitoli, etc..
 - Esempi: Quicktime di Apple, WebM di Google, AVI di Microsoft, MKV
- *Streaming video*: video visualizzato in tempo reale mentre viene scaricato da Internet (Netflix)

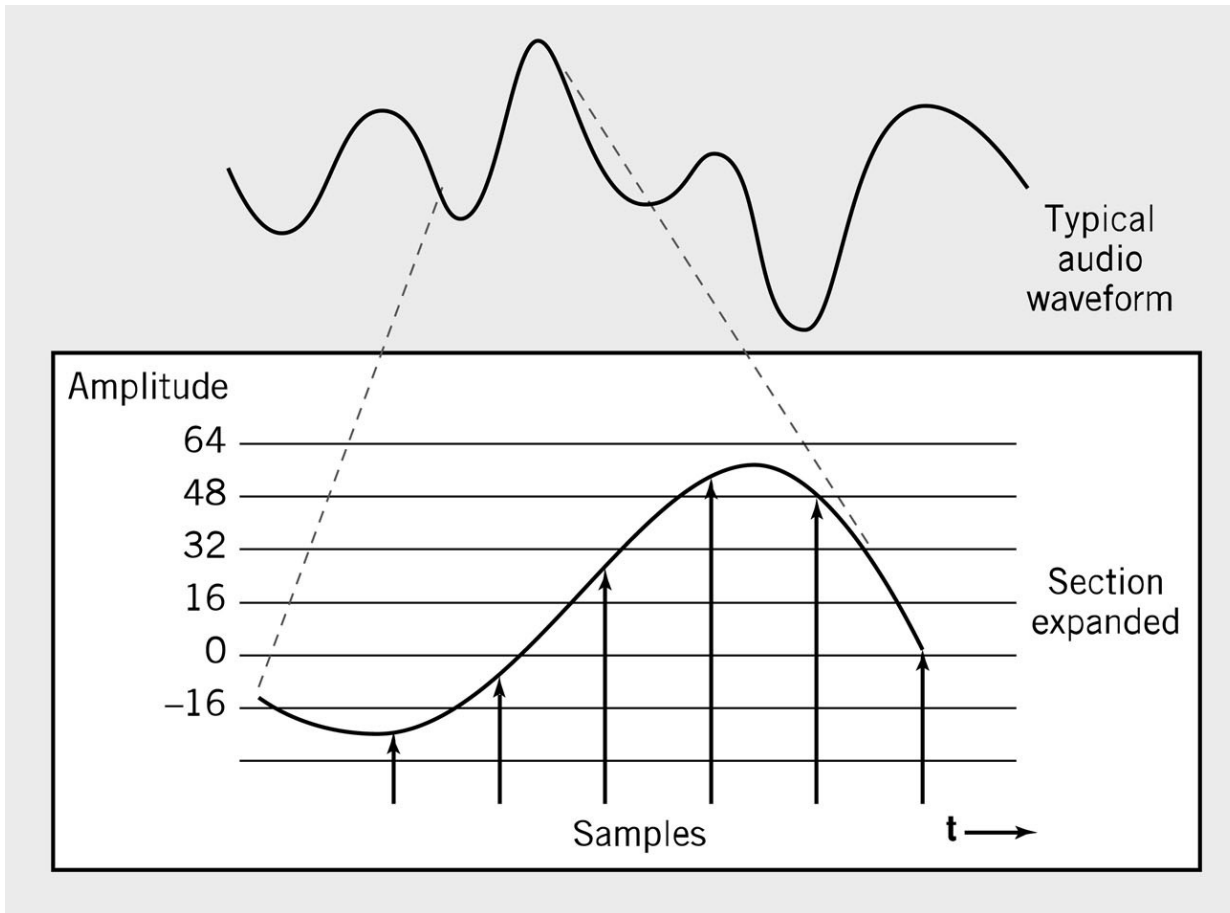


Audio

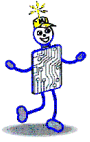
- La trasmissione ed elaborazione dell'audio è meno costosa di quella del video
- Forma d'onda analogica: rappresentazione digitale dell'audio
- Il suono analogico viene convertito in digitale da *convertitori analogico-digitali*
- **MIDI** (Musical Instrument Digital Interface): istruzioni per ricreare o sintetizzare suoni



Forma d'onda audio

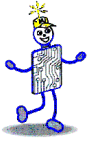


Tasso di campionamento normalmente di 44.1 o 48 KHz



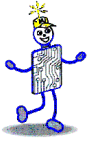
Frequenza di campionamento

- Numero di volte per secondo che il suono è misurato (campionato) durante la registrazione.
 - 1000 campioni al secondo = 1 KHz (kilohertz)
 - Esempio: CD Audio campionati as 44.1KHz
- Altezza di ogni campione memorizzata come:
 - Numero a 8 bit per qualità radiofonica
 - Numero a 16-bit per l'alta fedeltà
 - 2 x 16-bit per lo stereo

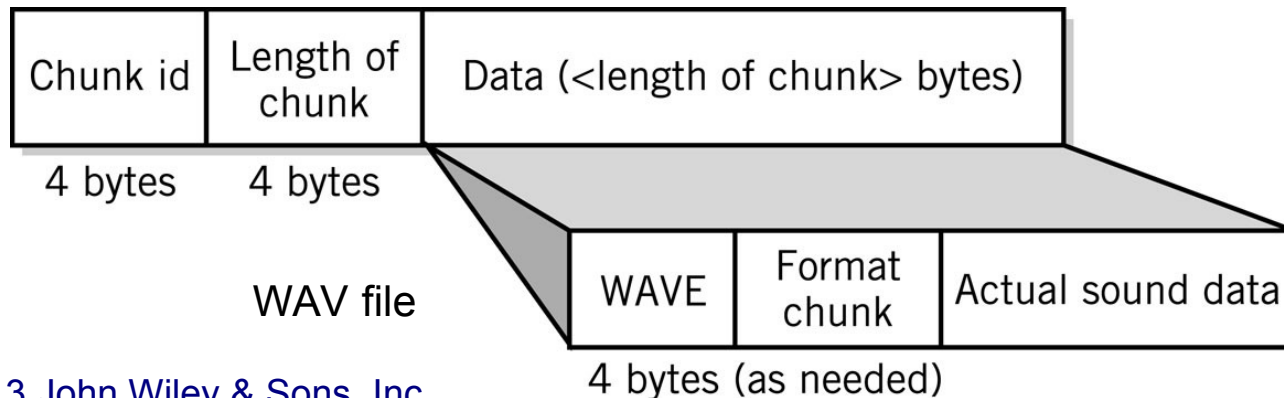
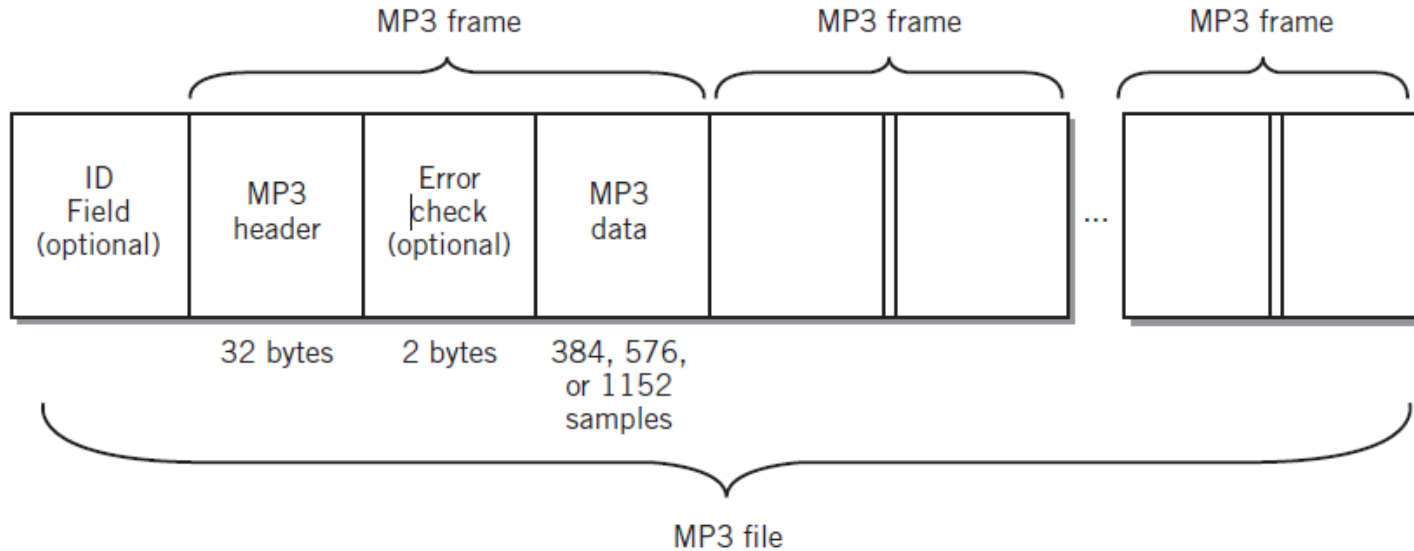


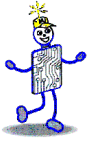
Formati audio

- *MP3* – formato audio digitale predominante
 - Derivato da MPEG-2 (ISO *M*oving *P*icture *E*xperts *G*roup)
 - Usa tecniche di compressione lossy psico-acustiche per ridurre lo spazio di immagazzinamento
- *WAV*
 - Sviluppato da Microsoft come parte delle sue specifiche multimediali
 - Formato di uso generale particolarmente adatto per piccoli frammenti di audio
 - Formato non compresso con 8 o 16 bit per campione (eventualmente stereo)



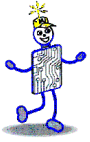
Audio Data Formats





Compressione dati

- *Compressione*: memorizzazione dei dati che riduce lo spazio di memoria occupato
- **Fattore di compressione**: rapporto tra la dimensione del file compresso e del file originario
- Compressione **lossless**: l'algoritmo di decompressione consente di ottenere esattamente il file originale:
 - Esempi: GIF, PCX, TIFF, ZIP
- Compressione **lossy**: consente una certa degradazione dei dati per un notevole miglioramento della occupazione di memoria
 - Fattori di compressione molto più alti, anche di 10 a 1
 - Esempi: JPEG, MP3
 - Comune per dati multimediali
- H.264: usa entrambi i tipi per raggiungere fattori di compressione di 10 a 1.



Linguaggi di descrizione pagina

- Descrivono la disposizione degli oggetti su una pagina da stampare o visualizzare sullo schermo
- Gli oggetti possono essere testo, immagini vettoriali, immagini bitmap, oggetti multimediali e altri tipi di dati
- Esempi
 - HTML, XHTML, XML
 - PDF
 - Postscript



Copyright 2013 John Wiley & Sons

All rights reserved. Reproduction or translation of this work beyond that permitted in section 117 of the 1976 United States Copyright Act without express permission of the copyright owner is unlawful. Request for further information should be addressed to the Permissions Department, John Wiley & Sons, Inc. The purchaser may make back-up copies for his/her own use only and not for distribution or resale. The Publisher assumes no responsibility for errors, omissions, or damages caused by the use of these programs or from the use of the information contained herein.”