Data Representation

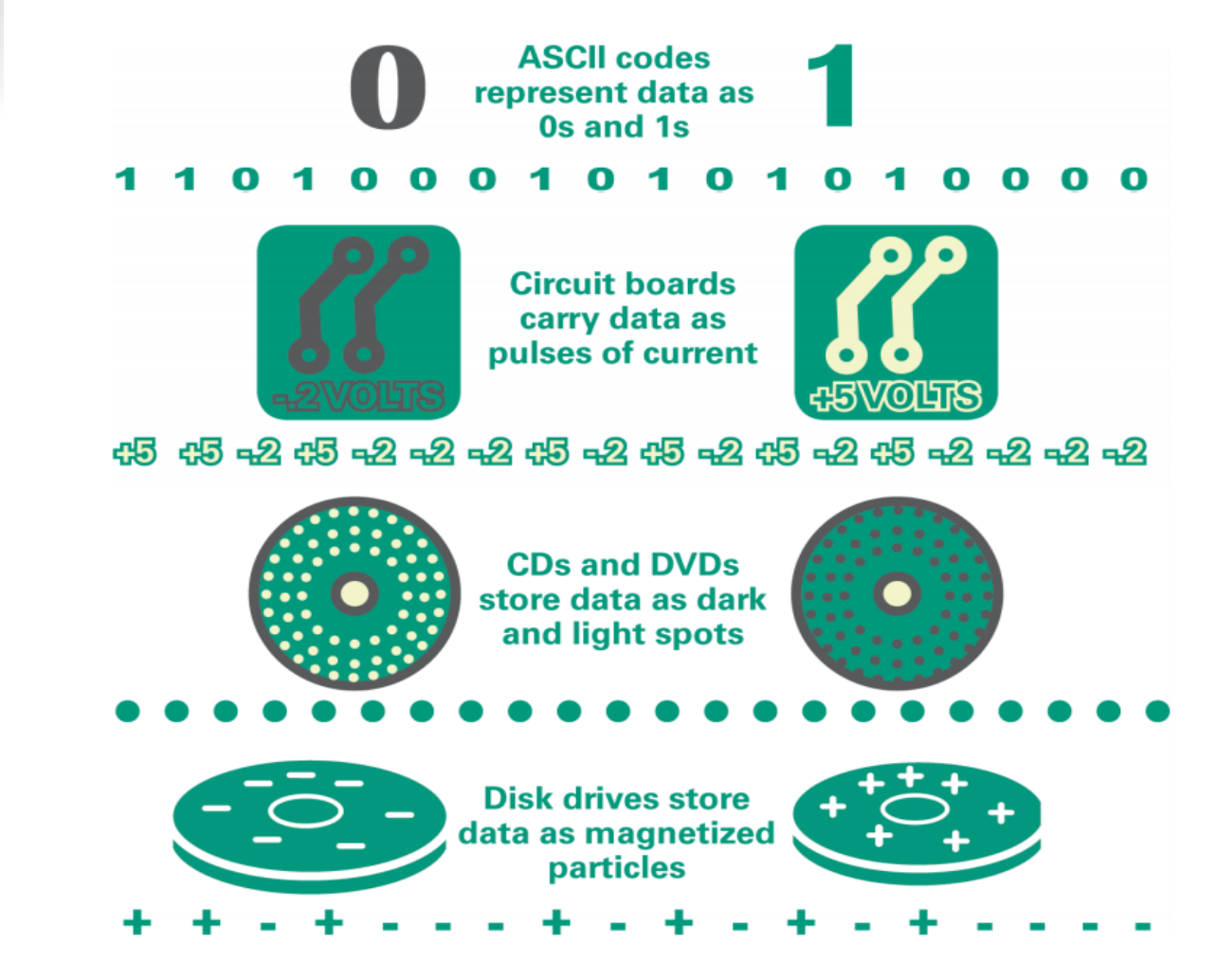
Data Representation

- **Data ~** refers to the symbols that represent people, events, things, and ideas.
  - Data can be a name, a number, the colors in a photograph, or the notes in a musical composition.
- **Data Representation ~** refers to the form in which data is stored, processed, and transmitted
- Devices such as smartphones, tablets, and computers store data in digital formats that can be handled by electronic circuitry.

Data Representation

- **Digitization ~** the process of converting information, such as text, numbers, photo, or music, into digital data
  - can be manipulated by electronic devices
- The 0s and 1s used to represent digital data are referred to as binary digits – the word **bit** stands for **B**inary dig**it**
- **digital file ~** or simply file is a named collection of data that exits on a storage medium
  - hard disk, CD, DVD, or flash drive

Data Representation



Representing numbers

- Numeric data consists of numbers that can be used in arithmetic operations.
- Digital devices represent numeric data using the binary number system, also called base 2.
- The binary number system only has two digits: 0 and 1.
- No numeral like 2 exists in the system, so the number “two” is represented in binary as 10 (pronounced “one zero”).

Representing Text

- Character data is composed of letters, symbols, and numerals that are not used in calculations.
  - Examples of character data include your name, address, and hobbies
- Character data is commonly referred to as “text.”

Representing Text – ASCII

- Digital devices employ several types of codes to represent character data, including ASCII, Unicode, and their variants.
- **ASCII** (American Standard Code for Information Interchange) requires seven bits for each character.
  - The ASCII code for an uppercase B is 1000010.

Representing Text – Extended ASCII

- **Extended ASCII** is a superset of ASCII that uses eight bits for each character.
  - For example, Extended ASCII represents the uppercase letter B as 01000010.
- Using eight bits instead of seven bits allows Extended ASCII to provide codes for 256 characters.

Representing Text – others

- **Unicode ~** uses sixteen bits and provides codes or 65,000 characters.
  - used for representing the alphabets of multiple languages (Japanese, Chinese, Arabic)
- **UTF-8 ~** a variable-length coding scheme that uses seven bits for common ASCII characters but uses sixteen-bit Unicode as necessary
  - used commonly to render web pages in Internet browsers

Representing Text

00100000	Space	00110011	3	01000110	F	01011001	Y	01101100	l
00100001	!	00110100	4	01000111	G	01011010	Z	01101101	m
00100010	"	00110101	5	01001000	H	01011011	[	01101110	n
00100011	#	00110110	6	01001001	I	01011100	\	01101111	o
00100100	$	00110111	7	01001010	J	01011101	]	01110000	p
00100101	%	00111000	8	01001011	K	01011110	^	01110001	q
00100110	&	00111001	9	01001100	L	01011111	_	01110010	r
00100111	'	00111010	:	01001101	M	01100000	`	01110011	s
00101000	(	00111011	;	01001110	N	01100001	a	01110100	t
00101001	)	00111100	<	01001111	O	01100010	b	01110101	u
00101010	*	00111101	=	01010000	P	01100011	c	01110110	v
00101011	+	00111110	>	01010001	Q	01100100	d	01110111	w
00101100	,	00111111	?	01010010	R	01100101	e	01111000	x
00101101	-	01000000	@	01010011	S	01100110	f	01111001	y
00101110	.	01000001	A	01010100	T	01100111	g	01111010	z
00101111	/	01000010	B	01010101	U	01101000	h	01111011	{
00110000	0	01000011	C	01010110	V	01101001	i	01111100	|
00110001	1	01000100	D	01010111	W	01101010	j	01111101	}
00110010	2	01000101	E	01011000	X	01101011	k	01111110	~

Representing Text

- ASCII codes are used for numerals
  - credit card numbers
  - phone numbers
- Plain, unformatted text is sometimes called ASCII text and is stored in a text file with a name ending in ".txt".
  - Windows devices label these files as "Text Document"
  - In Apple, these files are labeled "Plain Text"
- ASCII text files contain no formatting
  - creating documents with styles and formats, formatting codes have to be embedded in the text

Representing Text

- Microsoft Word produces formatted text and creates documents in **DOCX** format.
- Adobe Acrobat produces documents in **PDF** format.
- HTML markup language used for Web pages produces documents in **HTML** format.
- Apple Pages produces documents in **PAGES** format.

Representing Text

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<w:document xmlns:wpc="http://schemas.microsoft.com/office/word/2010/wordprocessingCanvas" xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006"
xmlns:o="urn:schemas-microsoft-com:office:office" xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships" xmlns:m="http://schemas.openxmlformats.org/
officeDocument/2006/math" xmlns:v="urn:schemas-microsoft-com:vml" xmlns:wp14="http://schemas.microsoft.com/office/word/2010/wordprocessingDrawing" xmlns:wp="http://schemas.
openxmlformats.org/drawingml/2006/wordprocessingDrawing" xmlns:w10="urn:schemas-microsoft-com:office:word" xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main"
xmlns:w14="http://schemas.microsoft.com/office/word/2010/wordml" xmlns:wpg="http://schemas.microsoft.com/office/word/2010/wordprocessingGroup" xmlns:wpi="http://schemas.microsoft.com/
office/word/2010/wordprocessingInk" xmlns:wne="http://schemas.microsoft.com/office/word/2006/wordml" xmlns:wps="http://schemas.microsoft.com/office/word/2010/wordprocessingShape"
mc:Ignorable="w14 wp14"><w:body><w:p w14:paraId="408A64C7" w14:textId="77777777" w:rsidR="004B4937" w:rsidRDefault="004B4937"><w:bookmarkStart w:id="0" w:name="_GoBack"/
><w:bookmarkEnd w:id="0"/></w:p><w:tbl><w:tblPr><w:tblStyle w:val="TableGrid"/><w:tblW w:w="10296" w:type="dxa"/><w:tblLayout w:type="fixed"/><w:tblLook w:val="04A0" w:firstRow="1"
w:lastRow="0" w:firstColumn="1" w:lastColumn="0" w:noHBand="0" w:noVBand="1"/></w:tblPr><w:tblGrid><w:gridCol w:w="1548"/><w:gridCol w:w="8748"/></w:tblGrid><w:tr w:rsidR="00D466CE"
w:rsidRPr="009B6231" w14:paraId="408A64CC" w14:textId="77777777" w:rsidTr="003824B3"><w:tc><w:tcPr><w:tcW w:w="1548" w:type="dxa"/></w:tcPr><w:p w14:paraId="408A64C8"
w14:textId="77777777" w:rsidR="00D466CE" w:rsidRPr="009B6231" w:rsidRDefault="00A04FD5" w:rsidP="00D466CE"><w:pPr><w:rPr><w:rFonts w:cs="Calibri"/><w:b/><w:sz w:val="24"/><w:szCs
w:val="24"/></w:rPr></w:pPr><w:r w:rsidRPr="009B6231"><w:rPr><w:rFonts w:cs="Calibri"/><w:b/><w:sz w:val="24"/><w:szCs w:val="24"/></w:rPr><w:t>Course</w:t></w:r></w:p></
w:tc><w:tc><w:tcPr><w:tcW w:w="8748" w:type="dxa"/></w:tcPr><w:p w14:paraId="408A64C9" w14:textId="77777777" w:rsidR="002E328D" w:rsidRDefault="002E328D"
w:rsidP="002E328D"><w:pPr><w:rPr><w:rFonts w:cs="Calibri"/></w:rPr></w:pPr><w:r w:rsidRPr="002E328D"><w:rPr><w:rFonts w:cs="Calibri"/></w:rPr><w:t>IS:237-43P</w:t></w:r><w:r
w:rsidR="00E02228" w:rsidRPr="002E328D"><w:rPr><w:rFonts w:cs="Calibri"/></w:rPr><w:t xml:space="preserve"> </w:t></w:r><w:hyperlink r:id="rId11" w:tgtFrame="_blank" w:tooltip="Click
for a course description and list of prerequisites if applicable. Opens in a new window." w:history="1"><w:r w:rsidR="00E02228" w:rsidRPr="002E328D"><w:rPr><w:rFonts w:cs="Calibri"/
></w:rPr><w:t>Fundamentals of Information Assurance/Security</w:t></w:r></w:hyperlink><w:r><w:rPr><w:rFonts w:cs="Calibri"/></w:rPr><w:t xml:space="preserve"> </w:t></w:r></w:p><w:p
w14:paraId="408A64CA" w14:textId="77777777" w:rsidR="002E328D" w:rsidRPr="002E328D" w:rsidRDefault="001E71F5" w:rsidP="002E328D"><w:pPr><w:rPr><w:rFonts w:cs="Calibri"/></w:rPr></
w:pPr><w:r w:rsidRPr="002E328D"><w:rPr><w:rFonts w:cs="Calibri"/></w:rPr><w:t>3/19</w:t></w:r><w:r w:rsidR="002A157C" w:rsidRPr="002E328D"><w:rPr><w:rFonts w:cs="Calibri"/></
w:rPr><w:t xml:space="preserve"> to </w:t></w:r><w:r w:rsidR="00D71235" w:rsidRPr="002E328D"><w:rPr><w:rFonts w:cs="Calibri"/></w:rPr><w:t>5/12</w:t></w:r></w:p><w:p
w14:paraId="408A64CB" w14:textId="77777777" w:rsidR="0061344E" w:rsidRPr="009B6231" w:rsidRDefault="001E71F5" w:rsidP="002E328D"><w:pPr><w:rPr><w:rFonts w:cs="Calibri"/><w:b/><w:sz
w:val="24"/><w:szCs w:val="24"/></w:rPr></w:pPr><w:r w:rsidRPr="002E328D"><w:rPr><w:rFonts w:cs="Calibri"/></w:rPr><w:t>W</w:t></w:r><w:r w:rsidR="002E328D"
w:rsidRPr="002E328D"><w:rPr><w:rFonts w:cs="Calibri"/></w:rPr><w:t>ednesday and Friday</w:t></w:r><w:r w:rsidR="00E02228" w:rsidRPr="002E328D"><w:rPr><w:rFonts w:cs="Calibri"/></
w:rPr><w:t xml:space="preserve">, </w:t></w:r><w:r w:rsidRPr="002E328D"><w:rPr><w:rFonts w:cs="Calibri"/></w:rPr><w:t>8:30a.m.-11:00a</w:t></w:r><w:r w:rsidR="00E02228"
w:rsidRPr="002E328D"><w:rPr><w:rFonts w:cs="Calibri"/></w:rPr><w:t>.m.</w:t></w:r><w:r w:rsidR="002E328D"><w:rPr><w:rFonts w:cs="Calibri"/></w:rPr><w:t xml:space="preserve"> </w:t></
w:r></w:p></w:tc></w:tr><w:tr w:rsidR="00E53A42" w:rsidRPr="009B6231" w14:paraId="408A64CF" w14:textId="77777777" w:rsidTr="003824B3"><w:tc><w:tcPr><w:tcW w:w="1548" w:type="dxa"/></
w:tcPr><w:p w14:paraId="408A64CD" w14:textId="77777777" w:rsidR="00E53A42" w:rsidRPr="009B6231" w:rsidRDefault="00E53A42" w:rsidP="00D466CE"><w:pPr><w:rPr><w:rFonts w:cs="Calibri"/
```

Contents of a .docx file

Bits and Bytes

- All of the data stored and transmitted by digital devices is encoded as bits.
- Terminology related to bits and bytes is extensively used to describe storage capacity and network access speed.
- The word bit, an abbreviation for binary digit, can be further abbreviated as a lowercase **b**.
- A group of eight bits is called a byte and is usually abbreviated as an uppercase **B**.

Bits and Bytes

- When reading about digital devices, you'll frequently encounter references such as 90 kilobits per second, 1.44 megabytes, 2.8 gigahertz, and 2 terabytes.
- **Kilo**, **mega**, **giga**, **tera**, and similar terms are used to quantify digital data.
- Use **bits** for data rates, such as Internet connection speeds, and movie download speeds.
- Use **bytes** for file sizes and storage capacities.

Bits and Bytes - usage

- **104 KB**: Kilobyte (KB or Kbyte) is often used when referring to the size of small computer files.
- **56 Kbps**: Kilobit (Kb or Kbit) can be used for slow data rates, such as a 56 Kbps (kilobits per second) dial-up connection.
- **50 Mbps**: Megabit (Mb or Mbit) is used for faster data rates, such as a 50 Mbps (megabits per second) Internet connection.
- **3.2 MB**: Megabyte (MB or MByte) is typically used when referring to the size of files containing photos and videos.
- **100 Gbit**: Gigabit (Gb or Gbit) is used for really fast network speeds.
- **1 TB**: Terrabyte (TB or TByte) is commonly used to refer to storage capacity.

Data Compression

- To reduce file size and transmission times, digital data can be compressed.
- **Data compression ~** refers to any technique that recodes the data in a file so that it contains fewer bits.
- Compression is commonly referred to as “zipping.”

Data Compression

- Compression techniques divided into two categories: lossless and lossy
  - **Lossless compression ~** provides a way to compress data and reconstitute it into its original state
    - uncompressed data stays exactly the same as the original data
  - **Lossy compression ~** throws away some of the original data during the compression process
    - uncompressed data is not exactly the same as the original

Data Compression

- Software for compressing data is sometimes referred to as a compression utility or a zip tool.
- On laptops and desktop computers, the compression utility is accessed from the same screen used to manage files.
  - Can also come from 3rd party applications.
  - Example Winzip, WinRar etc.

Data Compression

- The process of reconstituting files is called extracting or unzipping.
- Compressed files may end with a **.zip**, **.rar**, **.gz**, **.pkg**, or **.tar.gz**.

End of Topic