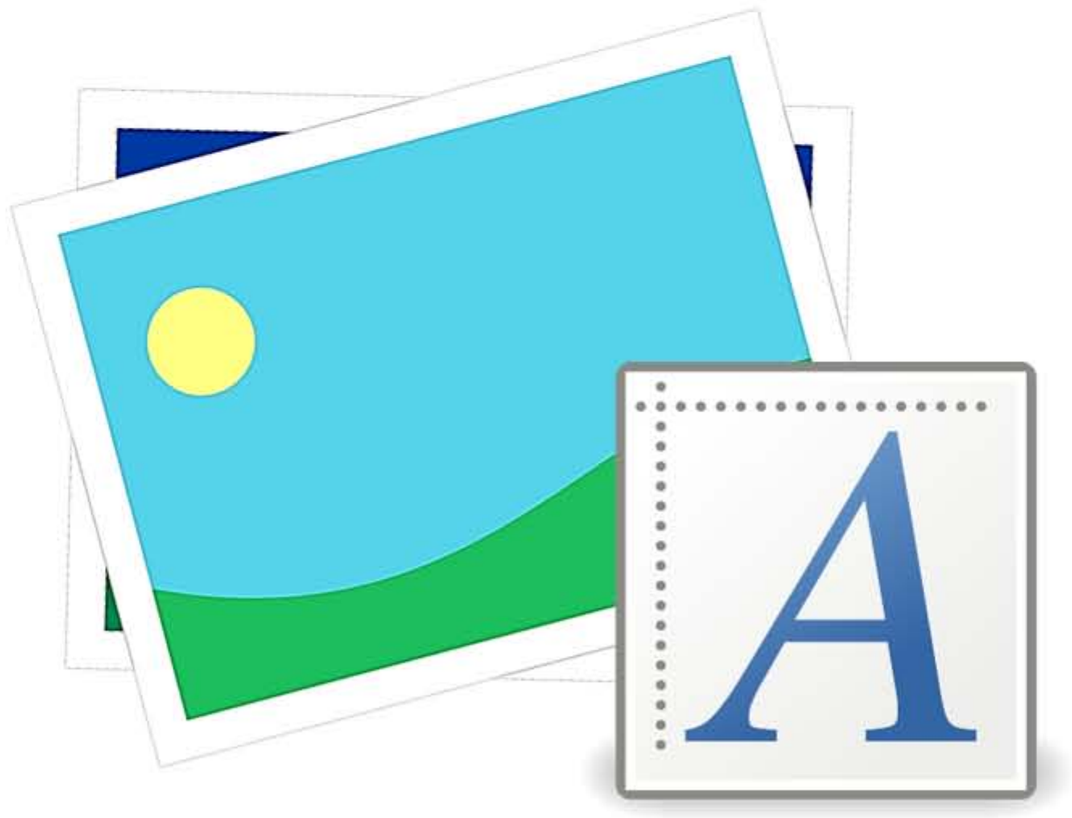
Data Representation and Compression



Characters

Computers can only represent data using binary digits. In order to represent letters and symbols we assign each character a code.

Character codes are numbers which computers can represent using binary.

ASCII

A common coding system is ASCII –
the **American Standard Code for Information Interchange**.

It uses seven bits for each character code. Seven bits is enough to code 128 different characters.

128 character codes is enough to represent the English alphabet plus a number of additional symbols, such as punctuation.

Character	Binary	Decimal
A	0100 0001	65
B	0100 0010	66
C	0100 0011	67
D	0100 0100	68
E	0100 0101	69
F	0100 0110	70

Compression

Compression is used to reduce the size of a file.

Compression can reduce the amount of storage space required and the time it takes to transfer the data over a network.

There are two types of compression:

Lossless

Lossy

Lossless Compression

Lossless compression reduces the size of a file without permanently removing any of the data.

This means the original data can be restored.

ZIP files are an example of lossless compression.

Lossy Compression

Lossy compression reduces the size of a file by permanently removing some of the data.

This means the original data cannot be restored.

Some examples of lossy compression algorithms are:

JPEG – used for images

MP3 – used for sound

MP3 compression works by removing frequencies that are out of the range of human hearing.