WHAT IS COMPRESSION?

Compression is the process of reducing the size of a file by encoding its data information more efficiently. By doing this, the result is a reduction in the number of bits and bytes used to store the information. In effect, a smaller file size is generated in order to achieve a faster transmission of electronic files and a smaller space required for its downloading.

#### HOW DOES COMPRESSION WORK?

When you have a file containing text, there can be repetitive single words, word combinations and phrases that use up storage space unproductively. Or there can be media such as high tech graphical images in it whose data information occupies too much space. To reduce this inefficiency electronically, you can compress the document.

Compression is done by using compression algorithms (formulae) that rearrange and reorganize data information so that it can be stored more economically. By encoding information, data can be stored using fewer bits. This is done by using a compression/decompression program that alters the structure of the data temporarily for transporting, reformatting, archiving, saving, etc.

Compression, when at work, reduces information by using different and more efficient ways of representing the information. Methods may include simply removing space characters, using a single character to identify a string of repeated characters, or substituting smaller bit sequences for recurring characters. Some compression algorithms delete information altogether to achieve a smaller file size. Depending on the algorithm used, files can be adequately or greatly reduced from its original size.

#### LOSSLESS COMPRESSION VS. LOSSY COMPRESSION

**Lossless Compression** is a type of compression that can reduce files without a loss of information in the process. The original file can be recreated exactly when uncompressed. To achieve this, algorithms create reference points (substitution characters) for things such as textual patterns, store them in a catalogue and send them along with the smaller encoded file. When uncompressed, the file is "re-generated" by using those documented reference points to re-substitute the original information.

Lossless compression is ideal for documents containing text and numerical data where any loss of textual information can't be tolerated. ZIP compression, for instance, is a Lossless compression that detects patterns and replaces them with a single character. Another example, LZW compression (Abraham Lempel, Jakob Ziv and Terry Welch-creators of LZW), works best for files containing lots of repetitive data.

**Lossy Compression,** on the other hand, reduces the size of a file by eliminating bits of information. It permanently deletes any unnecessary data. This compression is usually used with images, audio and graphics where a loss of quality is affordable. However, the original file can't be retained.

For instance, in an image containing a green landscape with a blue sky, all the different and slight shades of blue and green are eliminated with compression. The essential nature of the data isn't lost-the essential colours are still there. One popular example of Lossy compression is JPEG compression (Joint Photographic Experts Group) that is suitable for grayscale or colour images.