Quoted-Printable (from “An Internet Encyclopedia”)

The Quoted-Printable encoding is intended to represent data that largely consists of octets that correspond to printable characters in the ASCII character set. It encodes the data in such a way that the resulting octets are unlikely to be modified by mail transport. If the data being encoded are mostly ASCII text, the encoded form of the data remains largely recognizable by humans. A body which is entirely ASCII may also be encoded in Quoted-Printable to ensure the integrity of the data should the message pass through a character- translating, and/or line-wrapping gateway.

In this encoding, octets are to be represented as determined by the following rules:

Rule #1 (General 8-bit representation): Any octet, except those indicating a line break according to the newline convention of the canonical (standard) form of the data being encoded, may be represented by an "=" followed by a two digit hexadecimal representation of the octet's value. The digits of the hexadecimal alphabet, for this purpose, are "0123456789ABCDEF". Uppercase letters must be used when sending hexadecimal data, though a robust implementation may choose to recognize lowercase letters on receipt. Thus, for example, the value 12 (ASCII form feed) can be represented by "=0C", and the value 61 (ASCII EQUAL SIGN) can be represented by "=3D". Except when the following rules allow an alternative encoding, this rule is mandatory.

Rule #2 (Literal representation): Octets with decimal values of 33 through 60 inclusive, and 62 through 126, inclusive, MAY be represented as the ASCII characters which correspond to those octets (EXCLAMATION POINT through LESS THAN, and GREATER THAN through TILDE, respectively).

Rule #3: (White Space): Octets with values of 9 and 32 MAY be represented as ASCII TAB (HT) and SPACE characters, respectively, but MUST NOT be so represented at the end of an encoded line. Any TAB (HT) or SPACE characters on an encoded line MUST thus be followed on that line by a printable character. In particular, an "=" at the end of an encoded line, indicating a soft line break (see rule #5) may follow one or more TAB (HT) or SPACE characters. It follows that an octet with value 9 or 32 appearing at the end of an encoded line must be represented according to Rule #1. This rule is necessary because some MTAs (Message Transport Agents, programs which transport messages from one user to another, or perform a part of such transfers) are known to pad lines of text with SPACEs, and others are known to remove "white space" characters from the end of a line. Therefore, when decoding a Quoted-Printable body, any trailing white space on a line must be deleted, as it will necessarily have been added by intermediate transport agents.

Rule #4 (Line Breaks): A line break in a text body, independent of what its representation is following the canonical representation of the data being encoded, must be represented by a (RFC 822) line break, which is a CRLF sequence, in the Quoted-Printable encoding. Since the canonical representation of types other than text does not generally include the representation of line breaks, no hard line breaks (i.e., line breaks that are intended to be meaningful and to be displayed to the user) should occur in the quoted-printable encoding of such types. Of course, occurrences of "=0D", "=0A", "0A=0D" and "=0D=0A" will eventually be encountered. In general, however, base64 is preferred over quoted-printable for binary data.

Rule #5 (Soft Line Breaks): The Quoted-Printable encoding REQUIRES that encoded lines be no more than 76 characters long. If longer lines are to be encoded with the Quoted-Printable encoding, 'soft' line breaks must be used. An equal sign as the last character on an encoded line indicates such a non-significant ('soft') line break in the encoded text.