2. General Considerations

Textual encoding begins with a line comprising "-----BEGIN ", a label, and "-----", and ends with a line comprising "-----END ", a label, and "-----". Between these lines, or "encapsulation boundaries", are base64-encoded data according to Section 4 of [RFC4648]. (PEM referred to this data as the "encapsulated text portion".) Data before the encapsulation boundaries are permitted and parsers MUST NOT malfunction when processing such data. Furthermore, parsers SHOULD ignore whitespace and other non-base64 characters and MUST handle different newline conventions.

The type of data encoded is labeled depending on the type label in the "-----BEGIN " line (pre-encapsulation boundary). For example, the line may be "-----BEGIN CERTIFICATE-----" to indicate that the content is a PKIX certificate (see further below). Generators MUST put the same label on the "-----END " line (post-encapsulation boundary) as the corresponding "-----BEGIN " line. Labels are formally case-sensitive, uppercase, and comprised of zero or more characters; they do not contain consecutive spaces or hyphen-minuses, nor do they contain spaces or hyphen-minuses at either end. Parsers MAY disregard the label in the post-encapsulation boundary instead of signaling an error if there is a label mismatch: some extant implementations require the labels to match; others do not.

There is exactly one space character (SP) separating the "BEGIN" or "END" from the label. There are exactly five hyphen-minus (also known as dash) characters ("-") on both ends of the encapsulation boundaries, no more, no less.

The label type implies that the encoded data follows the specified syntax. Parsers MUST handle non-conforming data gracefully. However, not all parsers or generators prior to this Internet-Draft behave consistently. A conforming parser MAY interpret the contents as another label type, but ought to be aware of the security implications discussed in the Security Considerations section. The labels described in this document identify container formats that are not specific to any particular cryptographic algorithm, a property consistent with algorithm agility. These formats use the ASN.1 "AlgorithmIdentifier" structure as described in section 4.1.1.2 of [RFC5280].

Unlike legacy PEM encoding [RFC1421], OpenPGP ASCII armor, and the OpenSSH key file format, textual encoding does *not* define or permit headers to be encoded alongside the data. Empty space can appear between the pre-encapsulation boundary and the base64, but generators SHOULD NOT emit such any such spacing. (The provision for this empty

area is a throwback to PEM, which defined an "encapsulated header portion".)

Implementers need to be aware that extant parsers diverge considerably on the handling of whitespace. In this document, "whitespace" means any character or series of characters that represent horizontal or vertical space in typography. In US-ASCII, whitespace means HT (0x09), VT (0x0B), FF (0x0C), SP (0x20), CR (0x0D) and LF (0x0A); "blank" means HT and SP; lines are divided with CRLF, CR, or LF. The common ABNF production WSP is congruent with "blank"; a new production W is used for "whitespace". The ABNF in Section 3 is specific to US-ASCII. As these textual encodings can be used on many different systems as well as on long-term archival storage media such as paper or engravings, an implementer ought to use the spirit rather than the letter of the rules when generating or parsing these formats in environments that are not strictly limited to US-ASCII.

Most extant parsers ignore blanks at the ends of lines; blanks at the beginnings of lines or in the middle of the base64-encoded data are far less compatible. These observations are codified in Figure 1. The most lax parser implementations are not line-oriented at all, and will accept any mixture of whitespace outside of the encapsulation boundaries (see Figure 2). Such lax parsing may run the risk of accepting text that was not intended to be accepted in the first place (e.g., because the text was a snippet or sample).

Generators MUST wrap the base64 encoded lines so that each line consists of exactly 64 characters except for the final line which will encode the remainder of the data (within the 64 character line boundary), and MUST NOT emit extraneous whitespace. Parsers MAY handle other line sizes. These requirements are consistent with PEM [RFC1421].

Files MAY contain multiple textual encoding instances. This is used, for example, when a file contains several certificates. Whether the instances are ordered or unordered depends on the context.

3. ABNF

The ABNF [RFC5234] of the textual encoding is:

6. Textual Encoding of Certificate Revocation Lists

Certificate Revocation Lists (CRLs) are encoded using the "X509 CRL" label. The encoded data MUST be a BER (DER strongly preferred; see Appendix B) encoded ASN.1 "CertificateList" structure as described in Section 5 of [RFC5280].

```
-----BEGIN X509 CRL-----
MIIB9DCCAV8CAQEwCwYJKoZIhvcNAQEFMIIBCDEXMBUGA1UEChMOVmVyaVNpZ24s
IEluYy4xHzAdBgNVBAsTFlZlcmlTaWduIFRydXN0IE5ldHdvcmsxRjBEBgNVBAsT
PXd3dy52ZXJpc2lnbi5jb20vcmVwb3NpdG9yeS9SUEEgSW5jb3JwLiBieSBSZWYu
LExJQUIuTFREKGMpOTgxHjAcBgNVBAsTFVBlcnNvbmEgTm90IFZhbGlkYXRlZDEm
MCQGA1UECxMdRGlnaXRhbCBJRCBDbGFzcyAxIC0gTmV0c2NhcGUxGDAWBgNVBAMU
D1NpbW9uIEpvc2Vmc3NvbjEiMCAGCSqGSIb3DQEJARYTc2ltb25Aam9zZWZzc29u
Lm9yZxcNMDYxMjI3MDgwMjM0WhcNMDcwMjA3MDgwMjM1WjAjMCECEC4QNwPfRoWd
elUNpllhhTgXDTA2MTIyNzA4MDIzNFowCwYJKoZIhvcNAQEFA4GBAD0zX+J2hkcc
Nbrq1Dn5IKL8nXLgPGcHv1I/le1MNo9t1ohGQxB5HnFUkRPAY82fR6Epor4aHgVy
b+5y+neKN9Kn2mPF4iiun+a4o26CjJ0pArojCL1p8T0yyi9Xxvyc/ezaZ98HiIyP
c3DGMNR+oUmSjKZ0jIhAYmeLxaPHfQwR
-----END X509 CRL-----
```

Figure 8: CRL Example

Historically the label "CRL" has rarely been used. Today it is not common and many popular tools do not understand the label. Therefore, this document standardizes "X509 CRL" in order to promote interoperability and backwards-compatibility. Generators conforming to this document MUST generate "X509 CRL" labels and MUST NOT generate "CRL" labels. Parsers SHOULD NOT treat "CRL" as equivalent to "X509 CRL".

7. Textual Encoding of PKCS #10 Certification Request Syntax

PKCS #10 Certification Requests are encoded using the "CERTIFICATE REQUEST" label. The encoded data MUST be a BER (DER strongly preferred; see Appendix B) encoded ASN.1 "CertificationRequest" structure as described in [RFC2986].