
Subject: Re: More WIN/UNIX -> MAC transitions
Posted by [Martin Schultz](#) on Wed, 23 Aug 2000 07:00:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi Ben,

I am not in a mood to try (and I don't have a Mac anyway) but there may be a way to write a generic ASCII file reader. The main trick is to read binary into a BYTE array and test for CR (13B) and LF (10B) characters, then reform them as you go along. Something like:

```
;; find CR and LF characters
wcr = Where(barray EQ 13B, crcnt)
wlf = Where(barray EQ 10B, lfcnt)
;; convert all CRs to LFs unless you have both (DOS)
IF crcnt EQ lfcnt THEN BEGIN
    ;; complement of wcr (will be easier in IDL 5.4 ;-))
    wncr = Where(barray NE 13B, cnt)
    IF cnt GT 0 THEN barray = barray
ENDIF ELSE IF crcnt GT 0 THEN BEGIN
    barray = 10B
ENDELSE

;; convert byte array to string
;; (i leave this up to you as part of the EPA exam ;-)
```

Now, of course, you can get more and more sophisticated and implement buffered reading etc. or make it even an ASCII file converter (easy, once you know that all line ends are marked with LF only).

Here is a quick summary of what I learned about the line-end characters on the various OS's:

Unix : LF only (10B)
DOS/Win: CR/LF (13B, 10B)
Mac: CR only (13B)

Cheers,
Martin

Ben Tupper wrote:

```
>
> Hello,
>
> I have learned a couple of items worth sharing regarding our continuing
> transition from UNIX/WIN to MAC.
```

> Specifically, we have been wrestling with ASCII text data files
> generated from a CTD device embedded in a Windows OS environ. Of
> course the issue that surfaces is the carriage return/line feed (CRLF)
> that is used in one environment and not the others (and I can't remember
> which does what... it doesn't matter really). Our files have an
> complicated (messy) header followed by columnar data. Our interest is
> in the header.
>
> My hope is that by sharing what we've learned, someone will point us to
> an even better solution.
>
> We have come up with three solutions...
>
> (1) Load the Win/DOS text format files onto a UNIX machine and use the
> DOS2UNIX command:
>
> unix> dos2unix -ascii infilename outfilename
>
> This format is very clean for the MAC, too.
>
> (2) Load the Win/DOS text format file into a Mac editor, like BBEdit
> (www.barebones.com) and save in a Mac format.
>
> This method is easy, but we have a bazillion of these files and I'm a
> pretty poor show at piece work (too much daydreaming!)
>
> (3) Use IDL to read the file and handle the extra control characters
> internally.
> We are using this method now because the enduser doesn't want to mess
> around with exchanging files, etc. Since the header is relatively
> small, there is little performance loss. In a nutshell we introduced a
> test
> for the contents of the most recent line read. If the line is empty
> (meaning there was an extra linefeed or some-such-thing) then read the
> next line. Here's a snippet...
>
> .
> .
> .
>
> ReadF,U, Input, Format='(A)'
> If N_elements(Input) EQ 0 Then ReadF,U, Input, Format='(A)' ;check
> for blanks
> .
> .
> .
>
> So, it works, even if a bit brutish. The most important thing, from

> our standing, is that the enduser doesn't have to care a hoot about the
> format of the file.
>
> Ben
>
> --
> Ben Tupper
> Bigelow Laboratory for Ocean Science
> West Boothbay Harbor, Maine
> btupper@bigelow.org
> note: email address new as of 25JULY2000

--

```

[[ Dr. Martin Schultz  Max-Planck-Institut fuer Meteorologie
[[
[[      Bundesstr. 55, 20146 Hamburg
[[
[[      phone: +49 40 41173-308
[[
[[      fax:  +49 40 41173-298
[[
[[ martin.schultz@dkrz.de
[[
[[

```
