Interoperability Problems: Samba, Windows, Rsync and Unicode

A late night experience, thank good I got it fixed right away...

Imagine this setup:

    - A Samba file server (primary) and a Windows file server (secondary).
    - A DFS root which points to the Samba server (as primary) and the Windows server (as secondary, which maintains a replica of the primary).
An rsync client (cwrsync) is installed on the Windows server to maintain the DFS replica

Imagine the fact:

    - Files/directories which contain special chars (eg. umlauts) in their name copied onto the Samba server end up there correctly (eg. hÃ¼tte.doc)

Imagine the problem:

    - Files/directories which contain special chars in their named copied through rsync onto the Windows server end up there mangled (eg. hÃ?Â¼tte.doc)

The reason for this behavious is simple: cwrsync is actually nothing else than traditional rsync compiled as win32-binary using cygwin. Now rsync has one major drawback: it's not (yet) unicode aware, which means that special characters in file names are not properly converted.

Now there are two ways to fix this. Either replace the bundled cygwin library (cygwin1.dll) with another one which is unicode-aware. You find one at the UTF-8 cygwin project website.
Another possibility would be the use of an alternative tool (anything else than rsync) which is unicode-aware or the .NET based rsync port.

I choose to go the UTF-8 cygwin way, which did the trick for me.

I just downloaded the 1.5.21-1 version (I checked the bundled cygwin1.dll version through properties dialog in explorer first to make sure they match), moved the original cygwin1.dll away and replaced it by the download version instead.

By the next time I ran cwrsync my filenames would just look as supposed. Wheew, what a night...!