

# OIS Recommendations for File and Directory Names for Long-term Preservation and Access

Andrea Goethals, Harvard University Library

Last modified: 07/31/09

File and directory names can affect the long-term portability and usability of digital files. These names need to be able to 'play well' with diverse operating systems, file systems and applications and protocols that may interact with them, including:

- packaging and compression software, e.g. TAR
- web protocols , e.g. the URL specification
- common character sets, e.g. US-ASCII

These names are likely to be used within DRS reports and web interfaces as file identifiers. For this reason, the length of the values should not be overly long or they will negatively affect the readability of these reports and displays.

## Characters

A safe set of characters to use for file names and directories is the POSIX Portable Filename Character Set.

- The allowed characters are 0-9, a-z, A-Z, '.', '\_', and '-'. In other words:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z  
a b c d e f g h i j k l m n o p q r s t u v w x y z  
0 1 2 3 4 5 6 7 8 9 . \_ -

- Use a single '.' character to separate the file name prefix from the file extension. In the case of file compression formats used on archive file formats (e.g. TAR), the double extension format is acceptable. For example:
  - file.tar.gz
  - file.tar.Z
  - file.tar.bz2

## Length

Modern operating systems, file systems and applications can handle relatively long file and directory names - at least 128 characters. However, long file names cause display problems in on-line displays because of the amount of space they take up. For this reason very long file names are discouraged.

- Recommend a limit of 64 characters for both the supplied file name and owner supplied name
- Recommend a limit of 254 characters for the supplied file name and supplied directory name combined