Research Data Management: File Naming Guidelines

First and foremost:
- **Keep file names short but descriptive**
- **Be consistent with established conventions**

Recommended practices:

1. Denote dates in YYYYMMDD format
> **DO**: Use 20140403
> DON'T: Use 04032013
> **BECAUSE: Computers will sort YYYYMMDD in chronological order.**

2. Use unique identifier, e.g., project name or grant number
> **DO**: CHHM
> DON'T: Centre for Hip Health and Mobility
> **BECAUSE: It is good to keep it short to avoid side scrolling and column adjustment.**

3. Identify document content, e.g., questionnaire or grant proposal
> **DO**: FileNm_**Guidelines**_20140409_v01.docx
> DON'T: FileNm_20140409.docx
> **BECAUSE: It is easier to find relevant files.**

4. Use underscores or hyphens as delimiters; avoid space and special characters, e.g., &, *, #, etc.
> **DO**: FileNm_Guidelines_20140409_v01.docx
> DON'T: FileFm Guidelines 2014 04 09 v01.docx
> **BECAUSE: Special characters are difficult to read and may be handled differently by different systems.**

5. Keep track of document version either sequentially or with a unique date and time, e.g., v01, v02, 20140403_1800, etc.
> **DO**: FileNm_Guidelines_20140409_**v01**.docx
> DON'T: FileNm_Guidelines_20140409_**Review**.docx OR
>   FileNm_Guidelines_20140409_**Investigation**.docx
> **BECAUSE: Incorporating version number or date and time makes it easier to document the versions.**

6. Avoid complex folder hierarchies
> **DO**: F:/ **Env/LIBR/**DataMgmt_FileFormats_20140409_v01.docx
> DON'T: F:/**Environment/Library/Woodward/Data/Education/Materials/Draft/2014/04/-**DataMgmt_FileFormats_20140409_v01.docx
> **BECAUSE: Complex folder hierarchies require more effort for file navigation and saving. System back-ups may also take longer.**

For more information, see:
U.K. Data Archive. (n.d.). Format your data. Retrieved from https://www.ukdataservice.ac.uk/manage-data/format.aspx

UK Libraries

Research Data Management: File Formats

A file format is a way of encoding information within a computer file so that it can be recognized by an application and accessed. It is indicated by the file name extension (generally a period followed by three or four letters). This allows the computer to recognize the content type in a document, e.g., text, video, etc. Additionally, file formatting is important as this may affect whether the file contents are accessible following long-term storage.

Considerations:

1. Proprietary and non-proprietary (open) formats

Proprietary formats are limited by software patents, lack of format specification details, or built-in encryption to prevent open usage by the public. This results in requiring specific software provided by one vendor in order to use the proprietary format. In contrast, an open format is a file format that is freely available for everyone to use. Because the specifications are released, open-source developers can write software to utilize the file format in the case that a particular vendor no longer supports the file format. This increases the chances that technological developments do not make particular file formats obsolete.

2. Industry format adoption

In some cases, an industry may treat specific file formats as a de facto standard even if the formats are proprietary and rely on expensive software. In those cases, it may be more convenient to use the same proprietary file format.

3. Technical dependencies

Technical dependencies are the degree to which a particular format depends on particular hardware, operating system, or software and how these dependencies might influence future usage of the media. Using non-proprietary file formats may decrease the risk of technical obsolescence by removing the dependency on the underlying technology.

4. File quality and file size

Each file type such as text, images, or sound has many file formats available. File quality, the representation of the given item's characteristics, is a large part of the file format decision. Encoding that handles high resolution will be larger than lower quality file formats. However, the trade-off comes at the cost of storage space and convenience in disseminating the file to others.

Recommended File Formats			
Databases	XML, CSV	**E-Books**	EPUB
Images	JPG2000, PNG, TIFF	**Sounds**	MP3, FLAC
Spreadsheets	CSV	**Texts**	TXT, CSV, PDF/A, ASCII, UTF-8
Videos	MPG, MOV, AVI		

For more information, see:

Library of Congress. (2019, April 23). Sustainability of digital formats: Planning for Library of Congress collections. Retrieved from https://www.loc.gov/preservation/digital/formats/index.html

Virginia Tech. (2017, June 16). Digital Library and Archives: Recommended file formats. Retrieved from http://etd.vt.edu/howto/accept.html

This handout was adapted with permission from a document made available by the University of British Columbia Libraries (https://researchdata.library.ubc.ca/files/2019/01/FileFormat_Recommendations_20140417_v02.pdf).

Libraries