

For more information, please contact:

The Digital Repository of Ireland
Royal Irish Academy
19 Dawson St,
Dublin 2

web. www.dri.ie
email. dri@ria.ie
twitter. [@dri_ireland](https://twitter.com/dri_ireland)

DRI File Formats

DRI and Digital File Format Choices

As part of DRI's overall strategy to be a certified Trusted Digital Repository (TDR), DRI is developing policy guidelines to assist depositing institutions in their long-term digital preservation. These include recommendations regarding the file formats that will be accepted for ingestion into DRI.

File format choices are an extremely important component of any digital preservation strategy. As technology develops, digital formats can change quickly and old formats become unreadable. If they are proprietary formats, long-term support for them may not be guaranteed as companies disappear or simply stop supporting older versions of a particular format. For this reason, DRI's strategy is to recommend formats that are as sustainable as possible in the long term.

File Format Identification

File format is not necessarily indicated by file extension. While extensions are used by the operating system to associate file types with specific programs, they are not enforced. For example a .txt file can easily be renamed as a .doc file without making any changes to the file itself. It is important therefore, to identify file formats to ensure long term accessibility and preservation.

Video files are created using complex processes of compression and encoding known as codecs and then placed in

wrappers or containers. Containers must be in readable formats but it is the codec that determines the quality of the file. DRI recommends that codec file formats are checked before ingestion to the repository as the container name (which determines the file extension) does not necessarily determine the codec.

Some open source tools currently available for file format identification:

PRONOM is an on-line information system about data file formats and their supporting software products developed by the Digital Preservation Department of the UK National Archives. [1]

DROID is a software tool also developed by The UK National Archives to perform automated batch identification of file formats. It's a free software tool developed to help automatically profile a wide range of file formats. For example, it will tell you what versions you have, their age and size, and when they were last changed. [2]

FIDO is a command-line tool to identify the file formats of digital objects. It is designed for simple integration into automated work-flows. [3]

JHOVE developed by Harvard University Library and JSTOR, provides functions to perform format-specific identification, validation, and characterization of digital objects. [4]

FITS developed by Harvard University Library, 'identifies, validates and extracts technical metadata for a wide range of

DRI Factsheet Series No. 3:

DRI File Formats

file formats. It acts as a wrapper, invoking and managing the output from several other open source tools'. [5]

File Formats in DRI

The DRI policy for ingestion of file formats will be reviewed regularly to reflect changes in technology and data collections.

Preferred formats are recommended archival formats, best suited for long-term digital preservation; acceptable formats are commonly used, but not ideal for long-term digital archiving. They are often lower quality or proprietary. If you have a choice, we recommend

capturing in preferred formats.

DRI does not recommend converting files from an acceptable to a preferred format [for instance, from MP3 to WAV, or JPEG to TIFF], as this will not improve quality, and may in fact reduce it.

Important Note: It is important to stress that these are the formats DRI is currently recommending, and the range of ingestible formats will continue to grow over time. The most recent version of this factsheet will always be published on our website so please check www.dri.ie/publications for the most up to date version.

DRI PREFERRED AND ACCEPTABLE FORMATS (JUNE 2015)

Audio (Preferred): AIFF, FLAC, WAV, BWAV

Audio (Acceptable): MP3, MP2

Image (Preferred): TIFF, PNG, JPEG2000

Image (Acceptable): JPEG

Text (Preferred): PDF/A, txt, XML, plain text

Text (Acceptable): PDF, DOC, DOCx

Video (Preferred): Motion JPEG 2000 (ISO/IEC 15444-4 codec in .mj2 container), AVI (uncompressed or motion JPEG codec in .avi container), QuickTime Movie (uncompressed or motion JPEG codec in .mov container), Apple ProRes (codec in .mov container)

Video (Acceptable): MPEG4 (H.263 or H.264 codec in .mov, .avi or .m4v container), MPEG2 (H.262 in .mpg, .mpeg or .m2v container)

[1] PRONOM: <http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx> (accessed 17 June 2015)

[2] DROID: <http://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/file-profiling-tool-droid/> (accessed 17 June 2015)

[3] FIDO: <http://openpreservation.org/technology/products/fido/> (accessed 17 June 2015)

[4] JHOVE: <http://jhove.sourceforge.net/> (accessed 17 June 2015)

[5] FITS: <http://projects.iq.harvard.edu/fits> (accessed 18 June 2015)

