November 2022

apac Association of Performing Arts Collections

Digital Preservation Series

PART 4: File Formats - Case Studies

Prepared by APAC Digital Preservation Working Group

Table of Contents

UNIVERSITY OF BRISTOL THEATRE COLLECTION CASE STUDY

File Formats and The National Review of Live Art archive: Case study

In 2006 the University of Bristol Theatre Collection began work on a large-scale project funded by the Arts & Humanities Research Council to digitise a collection of over 1,650 video tapes that had been used to document performances from the National Review of Live Art (NRLA). The analogue video tapes were deteriorating and digitisation was required to preserve the content which was at risk of becoming lost.

Master Copies

The project selected the AVI (Audio Video Interleave) format for the storage of 'digital master' copies. Unlike the digitisation of still images, where large lossless TIFF (Tag Image File Format) and DNG (Digital Negative) format files together with smaller, compressed JPEG (Joint Photographic Experts Group) files for use as access copies had emerged by the mid-2000s as a widely shared common standard across the archives sector, the digitisation of audio-visual material had no well-established equivalent.

However, when it came to selecting file formats to use for the digitisation of the analogue video tapes the NRLA project took a similar archival approach to the digitisation of still images, prioritising uncompressed or losslessly compressed formats to try and preserve as much content as possible from the original analogue tapes. AVIs do not have to use compression during the capture process so the digital files created were 'lossless', maintaining the archival principle of preserving content. At the time, smaller compressed video files, although having the advantages of requiring a lot less storage space and being easier to playback, would nevertheless always be a bit like having a bad photocopy of the original document. And the project team was aware that future uses of the NRLA video archive would inevitably demand high resolution materials (such as re-edits or digital restoration), and these would be best served with uncompressed video data. As long as the data captured was the best quality that could be achieved at the time, these files could continue to be returned to, allowing the widest possibilities for future use – for exhibition, screenings, publication, web platforms, and could always be re-sized and re-formatted accordingly.

Although not perfect - in that the AVI had the drawback of being a proprietary format developed by Microsoft rather than 'open source, meaning files might be more at a risk in the future, particularly if Microsoft were to end up withdrawing their technical support for the format - it had been widely adopted, being compatible with Microsoft's operating systems (the dominant system in use across the University of Bristol and supported by the University IT department) open-source players such as VLC and non-linear editing software such as Adobe's 'Premiere' editing software package. An early project framework document suggested creating 'uncompressed video data deposited in JPEG2000 frames inside a MotionJPEG wrapper, accompanied by audio encoded in WAV format' (a theoretically better choice perhaps), but when the team came to test and develop digitisation workflows they struggled to open the JPEG2000 frames as videos as there was no software readily available to view them at the time. So AVI ended up being the pragmatic choice.

Another drawback was the size of the AVI files – one hour of analogue tape became roughly one hundred gigabytes of data, and it was likely that the project would create over twenty terabytes of uncompressed video files: a lot of space to find in 2006! By the time the project was complete following minimal restoration work on some files and the creation of access copies, around thirty-five terabytes of data ended up being transferred to LTO (Linear Tape Open) data tapes. These were later transferred to the University of Bristol's Research Data Storage Facility, from where they are currently being ingested into the Theatre Collection's new digital preservation system, Preservica

Access Copies

From the AVIs a set of compressed files using the MPEG-2 video file format (the second standard developed by the Moving Pictures Expert Group), which have the file extension .mpg, were then created as access copies. MPEG-2 was a common DVD-video standard, and each file was burnt to a DVD, forming the set of access copies that could be requested for viewing in the Theatre Collection's Reading Room and a second set of MPEG-2 copies kept on a standalone PC that could be readily accessed by staff. (An MPEG-2 copy of the relevant performance documentation was also sent on a DVD to each artist.)

With the rapid expansion of the internet and demand for streamed video content quickly becoming the norm during this period, in 2008 a follow-on project was developed to build a website that could host video content from the collection where artists had agreed to the documentation of their performances being published online. A new set of FLV (Flash Video) files were created for this purpose, as the FLV file format had been designed for streaming over the internet, was easy to embed into web sites and could be compressed to a small enough size to make them very easily accessible.

But after the end of the follow-on project in 2010, both the MPEG-2 and the Flash Video formats became increasingly outdated. With FLV files being criticised for operation and security issues, the website developed for streaming video content from the NRLA archive had to be taken offline (eventually, Adobe withdrew technical support for the format too, at the end of 2020). DVDs were also being superseded by internet streaming services, so by 2020 a new set of access copies using the more recent MPEG-4 file format (again, from the Motion Pictures Expert Group and which uses the file extension .mp4) had been created. MPEG-4 is a good quality, widely adopted format that can utilise the most widely used H264 codec (see below for more about codecs), for sharing video online and these are the copies that are currently made available for researchers visiting the Theatre Collection who wish to view video content from the NRLA archive. We are also working on once again publishing these online, where we have artists' permissions to do so.

With the MPEG-4 access copies in place, the earlier FLV files are being disposed of as they are surplus to requirements and FLV is not suitable as a preservation format. (Returning to the earlier analogy, with our new set of quality facsimile copies, the .mp4s, available for streaming, we can now dispose of the earlier inferior photocopies, the .flvs.) However, we have decided to retain the MPEG-2 copies for the time being as a 'belt and braces' back-up to the AVI files, until we are confident that all the AVIs have successfully been ingested into Preservica. Fifteen years have passed since the analogue video tapes were digitised, and the tapes will have deteriorated further during this time. If an AVI file won't play we may not be able to return to the original tape to make another digital copy: the MPEG-2 file would be the only copy we have.

Possible Future Migration of Formats

But this is not the end of the file format story for the NRLA video archive. AVIs, as noted above, are not an ideal preservation format and we are tentatively starting to discuss whether we should migrate the proprietary AVIs within Preservica to the open standards of the Matroska (MKV) format with the FFV1 (FF [Fast Forward] Video 1] codec, once the ingest of the AVIs has been completed.

Video files are complicated in having both a container or 'wrapper' which holds the data and metadata, and a codec used to organise this data – to encode or decode it. File extensions often refer to the container, and whilst some containers, such as .mpg for example, use a particular codec, others like .avi can make use of different codecs. (The AVI files in the NRLA collection, for example, largely use the v210 codec, where the source is analogue U-Matic and VHS video, which is not recognised by all video playing software. For the later MiniDV source tapes where the content is already in a compressed digital format, the codec is DV Video 'dvsd'.)

The MKV/FFV1 combination is starting to emerge as a commonly shared standard in use by archives – including by the British Film Institute - in much the same way as TIFF/DNG did for photographs. It uses lossless compression (it compresses the file size without removing data), which means that the files would take up about two-thirds of the digital storage space of the AVIs making them more efficient and cost-effective but without compromising on quality, and as a well-used open archive standard they would be less at risk of obsolescence.

NATIONAL THEATRE CASE STUDY

Understand your file formats

A free utility we have found useful for understanding the scope of our files at the NT is Filelist from Jam Software. Running it from within PowerShell creates a list of all files and subfolders from within a designated directory. This can be created with, or without, file checksums, along with additional functionalities. The .csv (comma separated values) output file can then be used with Excel's Power Queries to provide a closer look into a directory. This is done by splitting the comma separated values from the Filelist into designated columns which represent the directory's folder structure.

This gives us the ability to filter between folder levels, file extensions or view files ordered by size. This is incredibly useful when working with any number of files from within a directory, as Excel functionality can be used to pinpoint inconsistent file names, locate duplicates, or 'see' invisible system files which can also be picked up by the Filelist.

Path Level 1	Path Level 2	Path Level 3	Path Level 4	Path Level 5 (Production)	Path Level 6	File name
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Beaux Stratagem, The (2015)	RNT_PP_1_3_329	RNT_PP_1_3_329.pdf
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Beaux Stratagem, The (2015)	RNT_PP_2_3_334	RNT_PP_2_3_334.jpg
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Mandate, The (2004)	RNT_PP_2_4_259	RNT_PP_2_4_259_image-only-bw.tif
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	13 (2011)	RNT_PP_1_3_301	RNT_PP_1_3_301.pdf
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	13 (2011)	RNT_PP_2_3_295	RNT_PP_2_3_295.pdf
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Bed (1989)	RNT_PP_1_4_110	RNT_PP_1_4_110.pdf
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Bed (1989)	RNT_PP_2_4_107	RNT_PP_2_4_107.jpg
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Bent (1990)	RNT_PP_1_2_130	RNT_PP_1_2_130.pdf
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Bent (1990)	RNT_PP_2_2_110	RNT_PP_2_2_110.jpg
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Berlin (2009)	RNT_PP_2_2_410	RNT_PP_2_2_410.jpg
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Bopha! (1987)	RNT_PP_1_4_86	RNT_PP_1_4_86.pdf
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Bopha! (1987)	RNT_PP_2_4_85	RNT_PP_2_4_85_cropped.jpg
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Bopha! (1987)	RNT_PP_2_4_85	RNT_PP_2_4_85.jpg
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Brand (1978)	RNT_PP_1_3_23	RNT_PP_1_3_23.pdf
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Brand (1978)	RNT_PP_2_3_18	RNT_PP_2_3_18.jpg
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Brand (1978)	RNT_PP_2_3_18	RNT_PP_2_3_18_alternative.jpg
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Phedre (2009)	RNT_PP_1_2_294	RNT_PP_1_2_294.pdf
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Phedre (2009)	RNT_PP_2_2_265	RNT_PP_2_2_265.jpg
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Phedre (2009)	RNT_PP_2_2_265	RNT_PP_2_2_265_title_only.jpg
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Connections (2004)	RNT_PP_1_8_58	RNT_PP_1_8_58.pdf
ap_l_prom_03	STUDIO_ARCHIVE	ARCHIVE	NT Production Collection	Closer (1997)	RNT_PP_1_4_191	RNT_PP_1_4_191.pdf

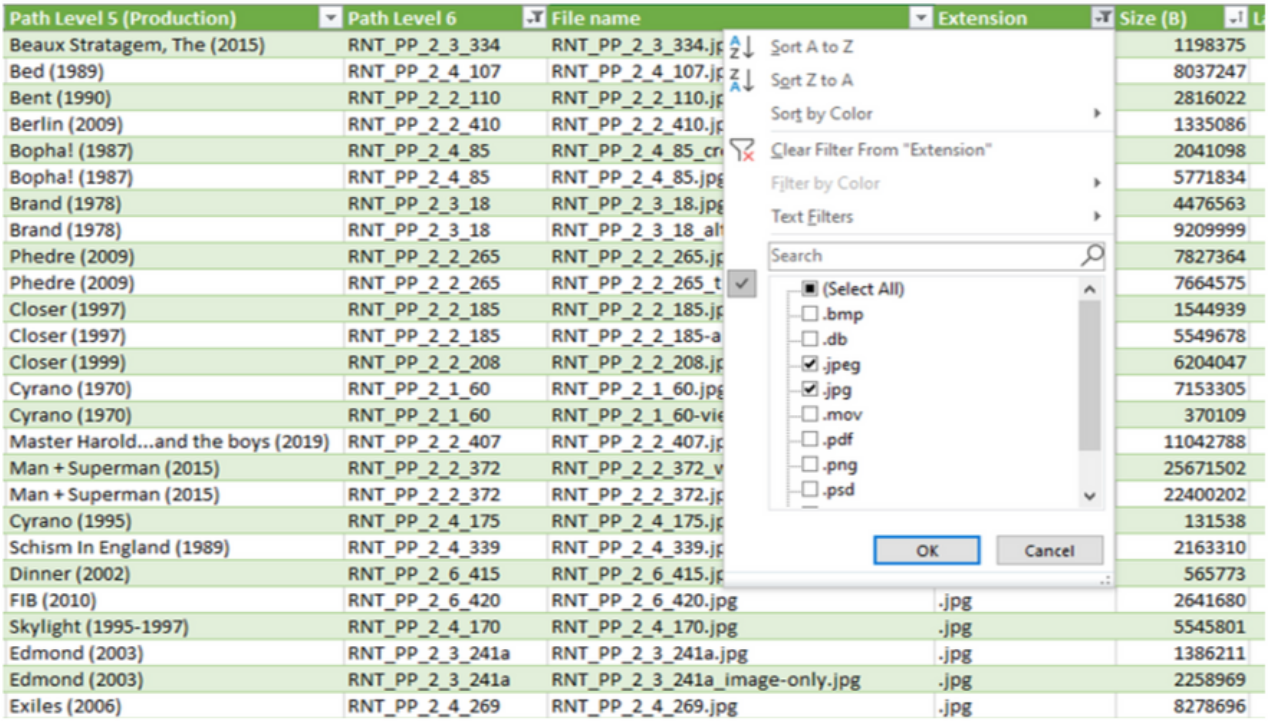
Figure 2.1: snapshot of .csv file created by Filelist

Filtering to show JPEG/JPG assets only

One particular issue that Filelist and Excel help us mitigate, is when certain 'illegal' characters crop up in file names (examples include: &?*$,) which can cause issues with our bulk ingests. We use customised Excel formulas to pinpoint file names which include any unaccepted characters and can then go on to change these. Overall, the combination of using Filelist and Excel has been incredibly helpful in managing the large number of files we hold in our Archive directory. The functionality that both utilities allow provides us with an in-depth snapshot of our digital files which we use for comparison and consistency.

Lists of files can be used to create a detailed Digital Asset Register. These documents were discussed in part 1 of this guidance and can be a useful starting point in your digital preservation journey.

At the NT we also use the PRONOM database, attached to Preservica. This database identifies file extensions and provides useful descriptions of how particular formats are used.

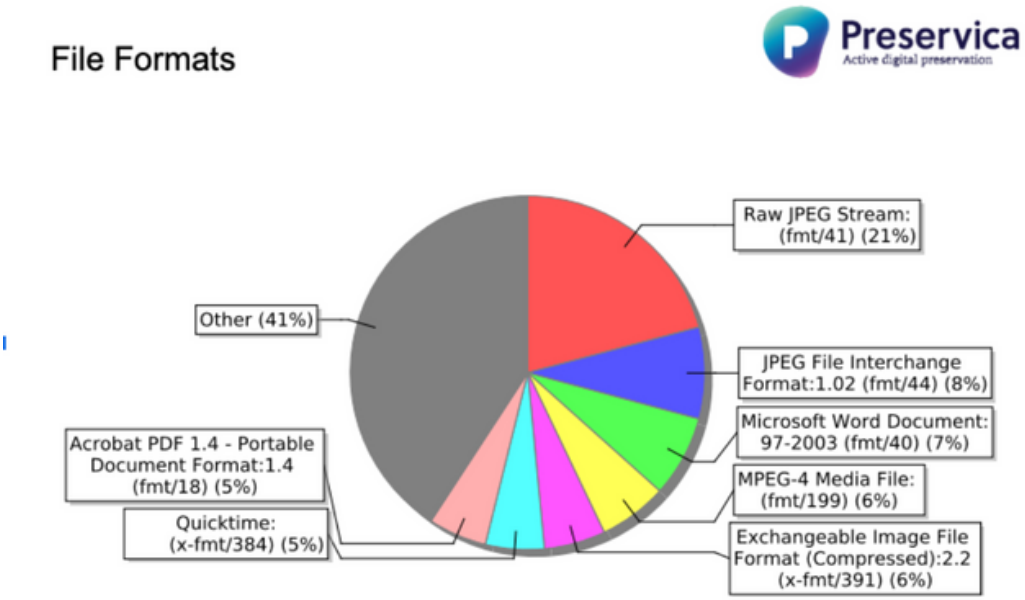


Path Level 5 (Production)	Path Level 6	File name	Extension	Size (B)
Beaux Stratagem, The (2015)	RNT_PP_2_3_334	RNT_PP_2_3_334.jp		1198375
Bed (1989)	RNT_PP_2_4_107	RNT_PP_2_4_107.jp		8037247
Bent (1990)	RNT_PP_2_2_110	RNT_PP_2_2_110.jp		2816022
Berlin (2009)	RNT_PP_2_2_410	RNT_PP_2_2_410.jp		1335086
Bopha! (1987)	RNT_PP_2_4_85	RNT_PP_2_4_85_cr		2041098
Bopha! (1987)	RNT_PP_2_4_85	RNT_PP_2_4_85.jpg		5771834
Brand (1978)	RNT_PP_2_3_18	RNT_PP_2_3_18.jpg		4476563
Brand (1978)	RNT_PP_2_3_18	RNT_PP_2_3_18_alt		9209999
Phedre (2009)	RNT_PP_2_2_265	RNT_PP_2_2_265.jp		7827364
Phedre (2009)	RNT_PP_2_2_265	RNT_PP_2_2_265_t		7664575
Closer (1997)	RNT_PP_2_2_185	RNT_PP_2_2_185.jp		1544939
Closer (1997)	RNT_PP_2_2_185	RNT_PP_2_2_185-a		5549678
Closer (1999)	RNT_PP_2_2_208	RNT_PP_2_2_208.jp		6204047
Cyrano (1970)	RNT_PP_2_1_60	RNT_PP_2_1_60.jpg		7153305
Cyrano (1970)	RNT_PP_2_1_60	RNT_PP_2_1_60-vie		370109
Master Harold...and the boys (2019)	RNT_PP_2_2_407	RNT_PP_2_2_407.jp		11042788
Man + Superman (2015)	RNT_PP_2_2_372	RNT_PP_2_2_372_v		25671502
Man + Superman (2015)	RNT_PP_2_2_372	RNT_PP_2_2_372.jp		22400202
Cyrano (1995)	RNT_PP_2_4_175	RNT_PP_2_4_175.jp		131538
Schism In England (1989)	RNT_PP_2_4_339	RNT_PP_2_4_339.jp		2163310
Dinner (2002)	RNT_PP_2_6_415	RNT_PP_2_6_415.jp		565773
FIB (2010)	RNT_PP_2_6_420	RNT_PP_2_6_420.jpg	.jpg	2641680
Skylight (1995-1997)	RNT_PP_2_4_170	RNT_PP_2_4_170.jpg	.jpg	5545801
Edmond (2003)	RNT_PP_2_3_241a	RNT_PP_2_3_241a.jpg	.jpg	1386211
Edmond (2003)	RNT_PP_2_3_241a	RNT_PP_2_3_241a_image-only.jpg	.jpg	2258969
Exiles (2006)	RNT_PP_2_4_269	RNT_PP_2_4_269.jpg	.jpg	8278696

Figure 2.2: snapshot of .csv file filtered to show jpeg/jpg assets only

Performing arts specific

The image below shows file formats ingested into National Theatre Preservica instance (August 2022)



File Format	Version	Format PUID	Frequency
Raw JPEG Stream		fmt/41	8463
JPEG File Interchange Format	1.02	fmt/44	3397
Microsoft Word Document	97-2003	fmt/40	3019
MPEG-4 Media File		fmt/199	2540
Exchangeable Image File Format (Compressed)	2.2	x-fmt/391	2275
Quicktime		x-fmt/384	2176
Acrobat PDF 1.4 - Portable Document Format	1.4	fmt/18	2167
MPEG 1/2 Audio Layer 3		fmt/134	1892
Microsoft Word for Windows	2007 onwards	fmt/412	1888
Acrobat PDF/A - Portable Document Format	1a	fmt/95	1880
JPEG File Interchange Format	1.01	fmt/43	1722
Broadcast WAVE	1 PCM Encoding	fmt/704	1678
Waveform Audio (PCMWAVEFORMAT)		fmt/141	1482
Acrobat PDF/A - Portable Document Format	2b	fmt/477	1336
Tagged Image File Format		fmt/353	815
Acrobat PDF 1.5 - Portable Document Format	1.5	fmt/19	729
Minolta RAW		fmt/669	368
Broadcast WAVE	0 PCM Encoding	fmt/703	307
Exchangeable Image File Format (Compressed)	2.2.1	fmt/645	274
Microsoft Excel 97 Workbook (xls)	8	fmt/61	228
RealLegal E-Transcript		fmt/728	217
Microsoft Outlook Email Message	97-2003	x-fmt/430	193
MIME Email	1.0	fmt/950	189
Microsoft Excel for Windows	2007 onwards	fmt/214	167
Material Exchange Format	Operational Pattern 1a	fmt/200	157
Acrobat PDF 1.3 - Portable Document Format	1.3	fmt/17	155
Acrobat PDF 1.7 - Portable Document Format	1.7	fmt/276	146
Acrobat PDF 1.6 - Portable Document Format	1.6	fmt/20	132
GZIP Format		x-fmt/266	53
Broadcast WAVE	0 Generic	fmt/1	52
Apple iWork Document	14	fmt/1441	46

Figure 2.3: File formats ingested into NT Preservica live instance (August 2022)

Our most common formats in the archive are image formats. We have a large volume of digitised and born-digital photographs. These are predominantly in standard formats such as JPEG and TIFF. Our file format report (produced using PRONOM) interestingly tells us that we have a number of unique JPEG formats in our repository, including a raw JPEG stream, the JPEG interchange file format 1.02 and 1.01.

As the National Theatre has many different departments, the archive is required to look after a wide range of formats covering the technical, cultural and administrative outputs of the theatre. Some of the most challenging areas are the records created by our production teams, including digital drawings.

We also advise departments on our file format preferences. For example, we have requested that our Press department now send us online press clippings in downloadable PDF formats, rather than as links, which can break or become inaccessible over time.

Moving image files have also proved complex. Some of our older .mov files are no longer renderable using standard players. We are also having to migrate some of our large master files to smaller .mp4 containers in order to make them renderable in our digital preservation software.

What to do with high risk assets

The NT Archive basement houses a cold store, which allows us to keep some materials at a cooler temperature than others e.g. VHS, hard drives, HD Cams, film reels, LTO tapes etc. This is the most basic level of preservation we undertake for digital content.

We maintain out of date technology when required to allow us to play old formats e.g., we have two reel to reel machines for tapes as well as VHS players. This allows us to play the content to identify it if required in order to make a decision about digitisation.

Through the work to develop a digital asset register, as discussed in the Part 1 Case Study, we identified some high risk assets, which were in proprietary formats. For example, the Digital Drawing team were using CAD files, which we cannot render in the research room for researchers. We asked the Digital Drawing team to provide a PDF-A version of the CAD files so that we could provide access to them in the research room. We still stored the CAD files in case the team needed to use them again in the future but we also ensured that there was a copy available for researcher access.

We have previously identified out of date software that has been in use across the NT and no longer supported by IT. One example was being used by the Press team to keep track of production data such as cast, creatives and awards won. The NT Archive team worked with the Press team to ensure that all of the data contained in their database was held elsewhere in the Archive and was accessible. The only information not held in the NT Archive were the awards so an export was done from the software and this was imported into the CALM performance module, allowing us to keep all of this data in a supported piece of software.

It was through identifying high risk assets and datasets that we were able to build a digital preservation policy and a solid case for a digital preservation solution.

CENTRE FOR CHINESE CONTEMPORARY ART CASE STUDY

Background

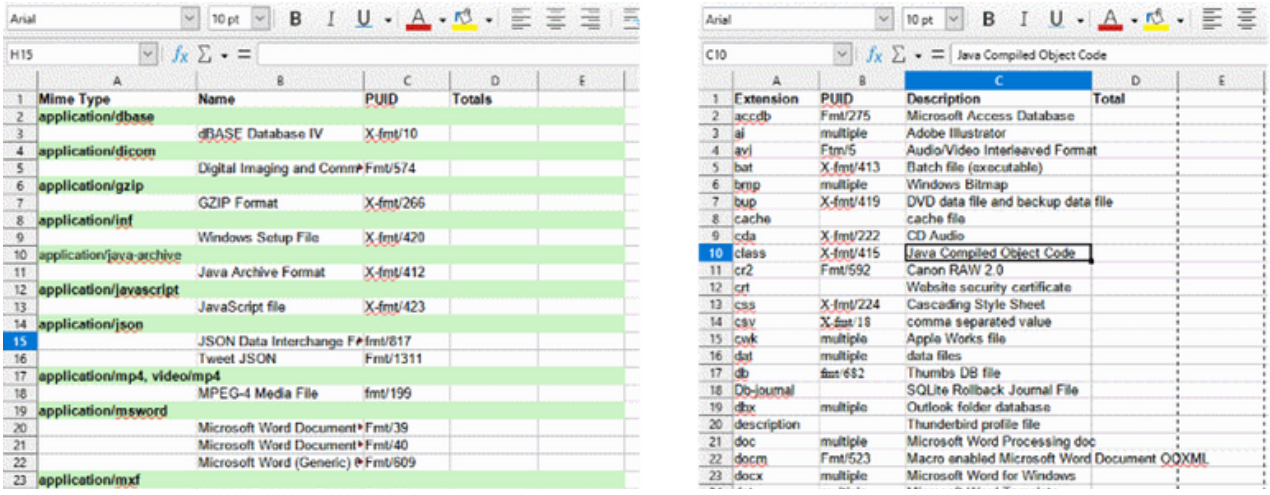
This case study looks at how we gather and process information on the file formats we hold within the records of the Centre for Chinese Contemporary Art (CFCCA). The archive collection comprises the documentation created by the Centre since 1986. The bulk of our archive collection dates from 1997 and around 70% of our current holdings are born-digital records. Over the last three years we have been gradually developing procedures to process our born-digital holdings with the resources we have available.

Identifying file formats

We collect data on our born-digital holdings using reports produced by the DROID software. The reports are created at the point materials are accessioned and to review our entire holdings. The original reports are exported as comma separated values (.csv) files. We also create copies of reports in Excel so they can be edited during our monitoring and documentation processes.

For new accessions, we used the report to check that all the file extensions are correct using the extension mismatch warning column. If possible, we create copies of the original files with the correct extension. We also review the file format count column to ensure that DROID has been able to correctly identify the files within the accession. In such cases we make sure the unidentified files can still be read and review their file extensions to try to identify the general file format.

We also use the information from the reports to maintain a master list of file formats. This spreadsheet contains information on the Mime Types, PRONOM PUID (Persistent Unique Identifier) and file extensions we have found within our collections. The information is used to speed up the process of creating more comprehensive reports and to keep the file format vocabularies in our database up to date.



	Mime Type	Name	PUID	Totals
2	application/dbase			
3		dBASE Database IV	X-fmt/10	
4	application/dicom			
5		Digital Imaging and Comm	Fmt/574	
6	application/gzip			
7		GZIP Format	X-fmt/266	
8	application/inf			
9		Windows Setup File	X-fmt/420	
10	application/java-archive			
11		Java Archive Format	X-fmt/412	
12	application/javascript			
13		JavaScript file	X-fmt/423	
14	application/json			
15		JSON Data Interchange F	fmt/817	
16		Tweet JSON	Fmt/1311	
17	application/mp4, video/mp4			
18		MPEG-4 Media File	fmt/199	
19	application/msword			
20		Microsoft Word Document	Fmt/39	
21		Microsoft Word Document	Fmt/40	
22		Microsoft Word (Generic)	Fmt/609	
23	application/mxf			

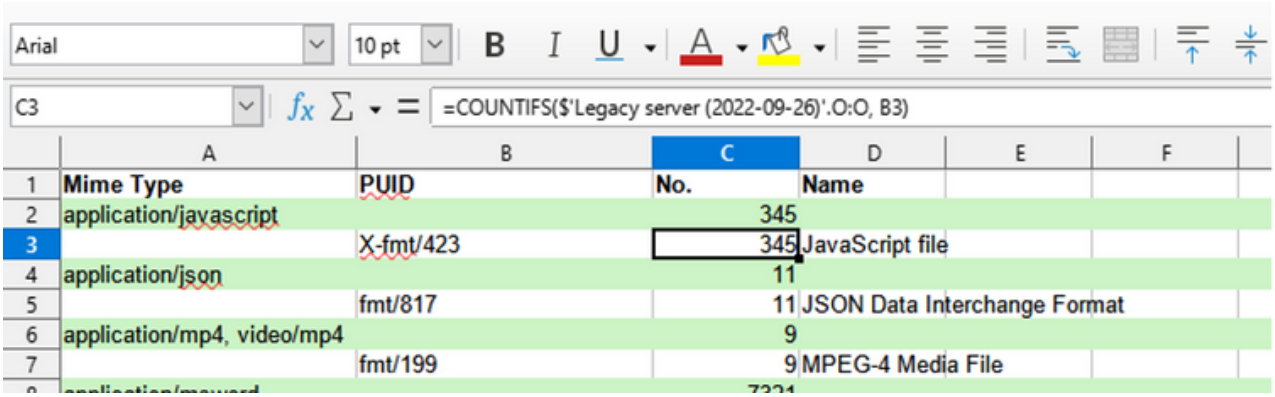
	Extension	PUID	Description	Total
2	accdb	Fmt/275	Microsoft Access Database	
3	ai	multiple	Adobe Illustrator	
4	avi	Fmt/5	Audio/Video Interleaved Format	
5	bat	X-fmt/413	Batch file (executable)	
6	bmp	multiple	Windows Bitmap	
7	bup	X-fmt/419	DVD data file and backup data file	
8	cache		cache file	
9	cda	X-fmt/222	CD Audio	
10	class	X-fmt/415	Java Compiled Object Code	
11	cr2	Fmt/592	Canon RAW 2.0	
12	crt		Website security certificate	
13	css	X-fmt/224	Cascading Style Sheet	
14	csv	X-fmt/18	comma separated value	
15	cwk	multiple	Apple Works file	
16	dat	multiple	data files	
17	db	fmt/682	Thumbs DB file	
18	Db-journal		SQLite Rollback Journal File	
19	dbx	multiple	Outlook folder database	
20	description		Thunderbird profile file	
21	doc	multiple	Microsoft Word Processing doc	
22	docm	Fmt/523	Macro enabled Microsoft Word Document OOXML	
23	docx	multiple	Microsoft Word for Windows	

Figure 3.1: View of the sheets in the master file format list

On a yearly basis we also use the DROID software to create a report to help us review all our digital holdings, including our semi-current records and our active file systems. This report is used to inform our cataloguing plans as well as help us plan for future digital preservation. We use DROID to review the contents of our archive server; our semi-current file systems (currently in 3 different folders: legacy, photo, and video), and the CFCCA active server.

Using the lists from the master file format spreadsheet, we create two additional sheets in the excel documents to collate data about file formats and file extensions. Using the COUNTIFS formula, we can count the number of times the format appears in the data produced by DROID.

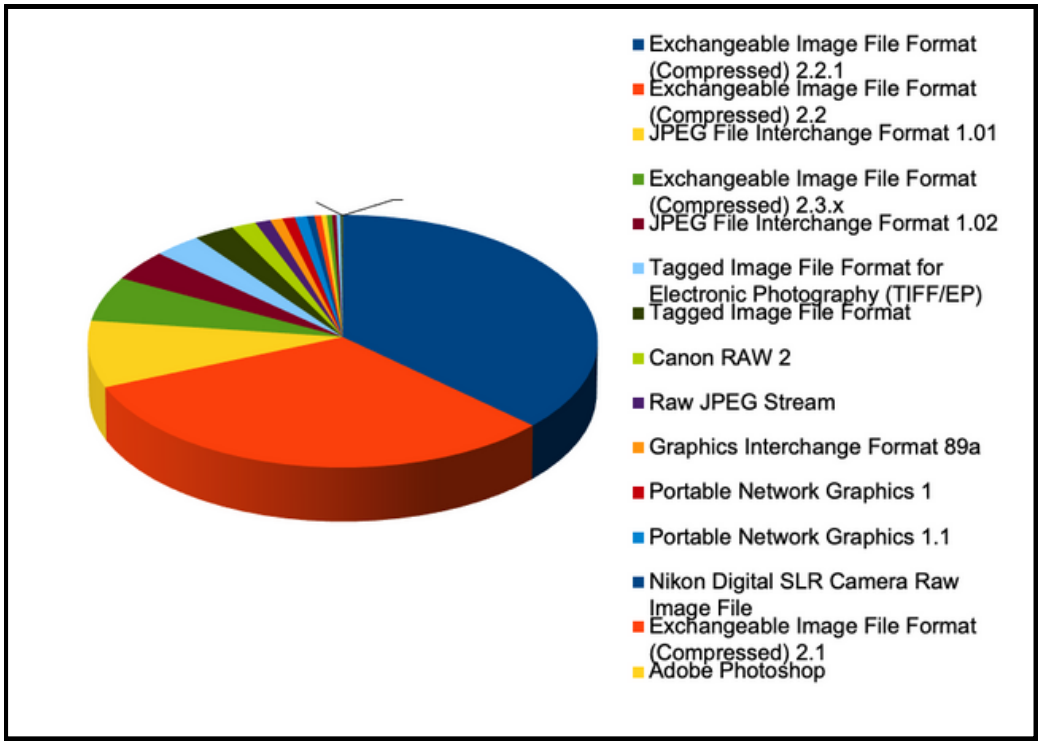
The data from the Excel documents is consolidated into a single spreadsheet which provide evidence for our future digital preservation programme.



	Mime Type	PUID	No.	Name
2	application/javascript		345	
3		X-fmt/423	345	JavaScript file
4	application/json		11	
5		fmt/817	11	JSON Data Interchange Format
6	application/mp4, video/mp4		9	
7		fmt/199	9	MPEG-4 Media File

Figure 3.2: COUNTIFS formula

We use this data to produce an overview of our born-digital records and other records which may be transferred to use in the future. Approximately around 70% of the files in our archival and semi-current holdings are standard images formats; 16% word processor documents; and 4% page layout files (such as pdfs). A further analysis of the types of image formats we have can be seen below:



Over 90% of the images we hold are jpeg formats, but also our collections hold a variety of file formats including tiffs, gif, png, bmps, and raw image formats.

Identifying high risk formats

Our born-digital holdings originate from back-up file servers and records held on physical digital media. Our priority is to make sure that all materials on physical digital media are back-up to our archive server to capture the records before the discs are no longer accessible. We create DROID reports for each item recording the contents of the media before and after it has been transferred to our server to carry out checksums to make sure the records have transferred correctly. Although we can still transfer the contents of CD/DVDs, we have found that some times DROID has difficultly reading the data from older media. In such cases, we capture what we can from the media and rely more on the report created from the backed-up files. We still have a back log of items to work through and this tends to be a background task which is carried out around other tasks.

Current we only have the faculties to migrate common Office documents, most image formats, and some video formats. We usually migrate records as part of our cataloguing processes, however we do try to prioritise more high risk formats such as video or older file formats. We can batch process image files, but other file formats have to be migrated individually. If we do not have software which can open a file format, we try to find an open source alternative we can use. There are file formats we hold which we can not process at this time, but these will be regularly check to ensure the aren't corrupted and our future preservation plans will take these into account.

Acknowledgements

Acknowledgement for the creation and development of this digital preservation series go to:

Erin Lee - National Theatre
Malcolm Mathieson - National Theatre
Arantza Barrutia-Wood - University of Sheffield
Bethany Johnstone - University College London
Robyn Greenwood - The Royal Shakespeare Company

With special thanks to:
Julian Warren - University of Bristol Theatre Collections
Hannah Smith - National Theatre
Katie Waring - Centre for Chinese Contemporary Art

and to the wider APAC digital preservation working group.

Stay tuned for
Part 5: Workflows & Mapping against NDSA Standards
to be published Winter 2022/23

Contact

info@performingartscollections.org.uk
@apac_ssn
performingartscollections.org.uk