ECP-2007-DILI-517006

EFG – The European Film Gateway

Guidelines for digitization, digital storage and retrieval

Deliverable number	*4.3*
Dissemination level	*Public*
Delivery date	*30 October 2009*
Status	*Final*
Author(s)	*Thomas C. Christensen (DFI), Julia Welter (DIF)*

e*Content*plus

This project is funded under the *e*Content*plus* programme,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

TABLE OF CONTENT

1 Introduction

1.1 Summary

For the purpose of making digitized film heritage available through the European Film Gateway and Europeana this report points at the most relevant file formats and means to create these formats. Especially within the area of moving images there are a plethora of formats and sub-formats, making it very difficult to make informed decisions. The aim of this report is to narrow down the confusion, and identify a pragmatic approach for the participating film archives.

This is necessary because:

a) some of the content to be made accessible described in the Description of Work, albeit already digitized, still rests on linear data storage on tape such as Digital Betacam. It has to be encoded and put into a non-linear hard disk environment before access via EFG and Europeana will be possible.

b) growth of moving image content in EFG beyond the collections described in the Description of Work is deemed necessary and desired. Hence, basic guidelines on how to encode and deploy moving images suitable for being displayed in EFG should be supplied to any institution beyond the current EFG consortium which intends to make moving image content available through EFG.

The European Film Gateway will support embedded display of moving image flash files, as well as JPEG files for still images. Other file types such as PDF, WAV, etc., may be listed in the EFG and linked to via the film archive's own web-pages, but will probably not be directly embedded in the EFG interface; here, an additional player or viewer will start (if installed on the client PC) when the user clicks on the item.

The file specifications for flash video are:

- 400x300 pixel, 500kbs (minimum)
- 768x572 pixel, 2Mbit/s (maximum)

The file specifications for JPEG still images are:

- 60.000 pixel (width x height) (minimum)
- 480.000 pixel (width x height) (maximum)

The file specifications, especially for moving images, are a pragmatic compromise between a minimum quality and offering content at a bitrate that gives an acceptable playback experience for most users.

1.2 Task description

Managing digitized content from film archives requires specific knowledge concerning content formats, storage and aspects of long-term stability and availability. Specific guidelines for film institutions will be compiled, to be distributed among the consortium and further film institutions. This will lay the foundation for the long-term availability of digital content via EFG. Since the question of formats is closely connected with the issue of players and bandwidth requirements, this task is fulfilled in close co-operation with DIF as the WP4 leader responsible for service implementation and web platform development.

1.3 Preface

This report presumes that selection of heritage has already taken place. In other words, this is a technical report, which describes the issues, consequences and possible dilemmas that arise in the archival handling of film related heritage. In a sense, this report will describe the HOW, but not the WHY or the WHAT. Though issues such as quality and quantity of digital content, and their cost consequences, are described, this report will try not to set up any answers as to which compromises are to be judged as better than others. However, a distinction between preservation grade quality and distribution grade quality will be made. The fact that digital distribution files and masters can be replaced by re-digitizing preservation elements, albeit sometimes at high cost, makes it possible to choose less costly conservation solutions for distribution files. Since items selected for preservation are considered artifacts of unique cultural value no compromises should be made in the preservation standards applied to these items.

Just as the disappearance and destruction of silent cinema in the early 1930s brought about the creation of the film archive movement, the digital revolution is creating a need and urgency to retain the analogue film heritage and to repurpose it in digital form. Film archives are determined to learn from the past, and to not allow technological change to destroy the cinema heritage of the 20th Century. Ironically, the new digital cinema contains some inherent, and largely unsolved, preservation challenges.

This report is moving image biased, since pictorial images, texts, and audio files, due to their much smaller file-size requirements, have largely been standardized both in regards to digital preservation formats and display formats. There is at this time no consensus on a single digital motion picture preservation or display format; such as PDF for text, TIFF for images and WAV for sound.

The purpose of the report is to enable the participating content providers in the European Film Gateway to make informed decisions on formats and file sizes for films to be disseminated online.

2 Preservation basics

The most reliable way to preserve the film heritage in the long term is to conserve the original film elements (the film negative) and to create duplicate film preservation masters. The effort to create backup copies is often referred to as film preservation. One might extend the term film preservation to also include efforts in restoration and presentation, which bring back the film to a cinema screen, or re-creates a viewing element resembling the original version. Preservation efforts for the film heritage therefore include conservation, duplication, restoration and presentation, as well as the preservation of film related materials such as posters, ads, stills and reviews.

For physical media, including digital carriers, the single most important factor in enhancing carrier longevity is a proper storage climate. The Image Permanence Institute has provided data and guidelines for the conservation and longevity of media carriers. Low temperature and relative humidity is essential to ensure the longevity of media carriers (Peter Adelstein (2004)).

Figure 1 is a guideline to mixed media storage. It is important to look at overall life expectancy for the media as well as to ensure the best storage. For instance film and analogue photos can achieve life expectancy in excess of 500 years, whereas magnetic tapes can only be expected to last up to 15 years. Also, digital media life expectancy is not only dependent on the media survival, but also the survival of play-back equipment and format obsolescence.

Carrier / Storage	Glass nega-tives	Ni-trate	Ace-tate B/W	Ace-tate Color	Poly-ester B/W	Poly-ester Color	Posi-tive photo B/W	Posi-tive photo Color	Inkjet	Mag-netic tape Ace-	Mag-netic tape Poly-	CD DVD
Room 20ºC												
Cool 12ºC												
Cold + 5ºC												
Frozen - 5ºC												

Figure 1. From Peter Adelstein (2004). Relative humidity: 30-50%.
Red indicates conditions that are likely to cause significant damage.
Yellow indicates conditions that do not meet ISO recommendations, but may be satisfactory for extended periods.
Green indicates conditions comparable to ISO recommendations.
Blue indicates conditions that will provide an extended lifetime.

It is worth mentioning that analogue media typically see a gradual signal/content deterioration, whereas digital media often either have signal or no signal. This means that some information typically can be recovered from analogue media once signal loss is detected, whereas digital media can be subject to instant and complete content loss.

For the material to be made available through the European Film Gateway, film transfer will typically have been made onto a Standard Definition (SD) master, such as Digital Betacam or an equivalent format. Further digitization into a file format has been performed for most content as well.

Digital preservation is an inevitable choice as more and more platforms turn digital, and more and more heritage is digitally born. However, commitment to a full digital preservation platform is a major endeavor, and storage cost alone should be expected within a Euro 500-5000 range per terabyte (TB) per year for a petabyte (PB) size digital repository. Apart from the storage cost, digital preservation includes creating an infrastructure to ensure the validity of the files ingested, as well as the maintenance of continuous migration of formats to new formats. Not only do the physical servers and storage media need to be upgraded frequently, but both the file formats and software operating systems are also subject to obsolescence. Failure to create a robust preservation infrastructure will increase the risk of data loss significantly. In digital preservation there is no room for the benign neglect that has sometimes been allowed in the analogue era.

Two schools of thought can be said to rule in the digital preservation field; one which believes it is best to migrate formats into new and current formats, and one that believes in preserving the original file formats, requiring access through emulation software running on the new platforms and operating systems.

EFG as an access platform is born into the migration practice, since one of the main objectives is to give as un-mediated access to the film collections included in the project. One of the main benefits of the “migration school” is that it limits the number of file types and only handles file types that are supported at the present time. The main concerns in regards to the “migration school” of thought are the risks involved in file conversion. File properties change as file formats evolve, and with each generation potential non-reversible changes may be incurred. It is worth noting that not only the file formats change, the operating systems also evolve; and this at an increasing speed.

Film archives have the opportunity to, and largely do, follow both schools of moving image preservation. Analogue film materials in good condition can be preserved without need for duplication for extended periods of time. The practice of film conservation in cool and cold climate conditions can be seen as preserving the original format to ensure authenticity and a direct link to the original, whereas the access is largely expected to be on other media. The

two schools of thought are therefore to be seen as complimentary, as are analogue and digital preservation.

2.1 Quality

Moving images can roughly be divided into three quality categories:

a) High resolution media (High Definition+ equivalent), theatrical presentation:
Bitrate: > 250Mbit/s
Analogue:
70/65mm film, 35mm film, 16mm film
Digital:
2K or 4K Digital Intermediate (dpx or equivalent), typical media: DTF, LTO
JPEG 2000 (DCDM/DCP)
Tape media: HDCAM SR, HDCAM, HDV
Blu-ray disc (mpeg4, mpeg2)

b) Standard Definition, or broadcast quality:
Bitrate: 50-250Mbit/s (preservation), 6-50Mbit/s (distribution)
Analogue:
Betacam SP, 1”C, U-Matic, many others
Digital:
Digital Betacam, DVCPro50 (DV 4:2:2 codec, 50Mbit/s), DVCAM, + many others
DVD disc (mpeg2)

c) Low-resolution (consumer and web applications, <6Mbit/s):
Bitrate: <6Mbit/s
Analogue:
VHS, Betamax
Digital:
DVD disc (mpeg2)
VCD disc (mpeg1)
Web-files: mpeg4/Ipod, flash, wmv, etc.

The categories are somewhat overlapping and a preservation master in SD can easily exceed an HD distribution file in both size and bit-rate. Although the web infrastructure and the average speed of internet connections in most households have improved significantly over the past few years, the “low resolution” solution is still the most prominent and widely used on the Internet.

Youtube for example uses the following specifications for the source file (from www.youtube.com):

Video Format: H.264, MPEG-2 or MPEG-4 preferred

- Aspect Ratio: Native aspect ratio without letterboxing (examples: 4:3, 16:9)
- Resolution: 640x360 (16:9) or 480x360 (4:3) recommended
- Audio Format: MP3 or AAC preferred
- Frames per second: 30
- Maximum length: 10 minutes (we recommend 2-3 minutes)
- Maximum file size: 1 GB

The stream-files used by Youtube are typically 320x240 (approximately 200 kbs) or 480x360 ("HD"). The proposed EFG stream files are therefore higher than the ones provided through Youtube.

3 Digitization

3.1 Moving images

For film heritage, preservation grade files are generally considered uncompressed JPEG2000 files in 4K (35mm), 2K (16mm), 16 bit in each color. These files are very large (4TB and 1TB per hour content) and are generally not yet considered relevant for preservation due to the cost of maintaining these files. Compared to the preservation of the original or re-recorded moving images on 35mm film the cost of preserving the equivalent digital content is a factor 10 to 100 more costly. In *The Digital Dilemma* (Milt Shefter & Andy Maltz, 2007), p. 1-2, it is concluded: "that the annual cost of preserving film archival master material is $1,059 per title, and the annual cost of preserving a 4K digital master is $12,514, an 11-fold difference." I his article *The Digital Black Hole*, Jonas Palm (2004) is concerned with the overall viability of digital storage systems. He warns that one should not expect digital preservation to be cheap, and concludes: "Whatever strategy one chooses to follow, the essential point to consider before undertaking large-scale digitization is the level of long-term financial commitment that can realistically be secured and to develop a preservation strategy accordingly. Estimations of costs that cover all aspects should be part of the planning process to limit the risk that a project ends up as yet another digital black hole, as so many others have done." (p.14).

Since the purpose of the EFG is to provide access to low resolution files for web display, the most relevant route is to give some pointers to medium term digitization at broadcast quality, and how to convert these media to the relevant digital files for medium term use and web purposes.

Film to broadcast transfer is a costly process, which is typically performed on film scanners or telecine workstations. Some archives have in-house facilities, but many outsource this work to specialized facility houses or laboratories. The typical master formats are HDSR, HDCAM or Digital Betacam video tapes. These tape formats are currently the best carriers for HD or SD quality, even if they have some inherent deficiencies such as being of limited physical stability (10-15 years life expectancy stored in 8C/40%RH), and being dependant on proprietary playback equipment. Runners up are DVCPro50, adopted by many European broadcasters, as well as the file formats JPEG2000, MPEG2 and MPEG4 AVC (H.264). However, the volume of digital content handled and owned in most film archives is still so small that a commitment to a parallel file based digital workflow at full broadcast quality is to most not yet a cost efficient investment, when compared with the physical media, which are compatible with the existing archival infrastructure.

The delivery format for EFG is currently foreseen to be Flash video; 640x480, 1000 kbs, approximately. This quality is slightly higher than known from Youtube, but not sufficient for other use than web applications. A watermark may or may not be applied by the supplying archive, depending on policies and IPR issues. Watermarks may be generated, both in the encoding procedure, and as part of the web-solution. Since the quality requirements for web publication will undoubtedly change in the short to medium term, an intermediate digital format is proposed to enable future batch conversions to other web display formats. For instance, some might find it relevant to set up podcasting, for which these intermediate files could easily be converted. The proposed formats for intermediate “master” use are MPEG2 (DVD compatible, 6 Mbit/s) for SD and MPEG2 or MPEG4 (Blu-ray compatible, 25-36 Mbit/s) for HD. To some it might make sense to look at JPEG2000, DVCPro50 (4:2:2) or MPEG2 (50Mbit/s 4:2:2), if finances and logistics apply. The reason for the choices involving Blu-ray and DVD compatible files as intermediate “masters” is that archives are often still engaged in giving access at higher resolutions (individual research or publication), for which these formats are standardized and practical. There is still a frequent request for physical DVD distribution to researchers, so the file formats compatible with physical disc media can provide synergy to archives to which this is relevant.

As the European Film Gateway deals with materials that are already digitized, the main issue is not how to do a proper analogue to digital conversion, but rather to bring the existing digital entities into a format that can easily be displayed on the EFG website.

Films should be made available from a local archive controlled webserver and service that allows deep links to the resource in a stream format. This requires a standardized setup that is compatible with the solution developed for the EFG website. Solutions that are currently in place work with wmv and flash stream servers. For partners that are not in a position to establish their own stream server for delivery of their films, a service option to be hosted by reelport will be established (see chapter 5).

The EFG underlying content, as described in the DoW section 4.1, is in a range of formats. The file based content is already in wmv or flash format, or can easily be converted into flash files. Some content is however in tape based form, which needs first to be converted into files. The physical format most mentioned is Digital Betacam, which is a high quality broadcast grade SD format. The partners with physical tape or disc formats will need to convert these media into pure file formats. Here too, reelport is able to offer support.

The bit rate of the stream files will be in a range of 300kbs to 2Mbit/s. The specific solution will be dependant on the choices of the individual content provider. However, a digital master file with a format and specification that exceeds this is recommended to ensure the medium term availability of the digitized content.

Digital Betacam has a bit rate of more than 100Mbit/s and is to many the preferred medium for storage and preservation of Standard Definition video. Digital preservation files can be created from Digital Betacam, and should be 4:2:2 with a minimum bit rate of 50Mbit/s. Preservation formats include JPEG2000, uncompressed video, mpeg2, mpeg4 and DVDPro50.

For the purpose of the European Film Gateway project, there is no need to create preservation grade digital files. It is sufficient to generate files that can easily be converted into the relevant stream format. A medium term digital master could be a DVD compatible MPEG2 file (6Mbit/s), or a number of other formats (i.e. MPEG4, avi, vob, mov) of similar quality that the individual content provider feels is most relevant to the specific infrastructure. This is not the least dependant on the level of the existing expertise of the content provider, and whether the archive is Windows or Mac based. It is however very relevant to create a Standard Definition file of at least DVD quality (6Mbit/s+), since it will ensure the availability of the content as bandwidth and quality requirements increase.

The following file formats are mentioned by the project partners in the DoW:

- MPEG2 (>2000 items)
- WMV (> 60.000 items)
- FLV (>300 items)
- AVI (>60.000 items)
- MPEG1 (3000 items)
- MPEG4
- TIFF (images)
- JPEG (images)
- PDF (text/images)
- WAV (sound)

The following physical formats are mentioned are mentioned by the project partners in the DoW:

- HDSR (<50 items)
- Digital Betacam (approx. 2500 items)

3.2 Digitization Workflow

3.2.1 From physical tape carrier (<3000 items):

Digital Betacam → video out → video capture → file → encoder → stream file

Converting carrier-based video formats (HDSR, Digital Betacam, DVD, VHS, etc.) requires the conversion of the video-stream/-output into data file form. This is done by a video capture device, either a card inserted in a computer, or a stand alone box/converter. Most video capture cards can handle a wide range of input and deliver a wide range of different files and settings. This report suggests creating masters in the MPEG family, either MPEG2 or MPEG4. Depending on the possibilities, capabilities, and strategic planning of the individual archive it is possible to either just create a 6Mbit/s distribution master, and to take the step to preservation grade, requiring a capture card capable of producing 50Mbit/s+ (4:2:2 or 4:4:4) files, as well as a data storage environment capable of handling and preserving 30GB+ data per hour content.

3.2.2 File based (software) conversion (>60.000 items):

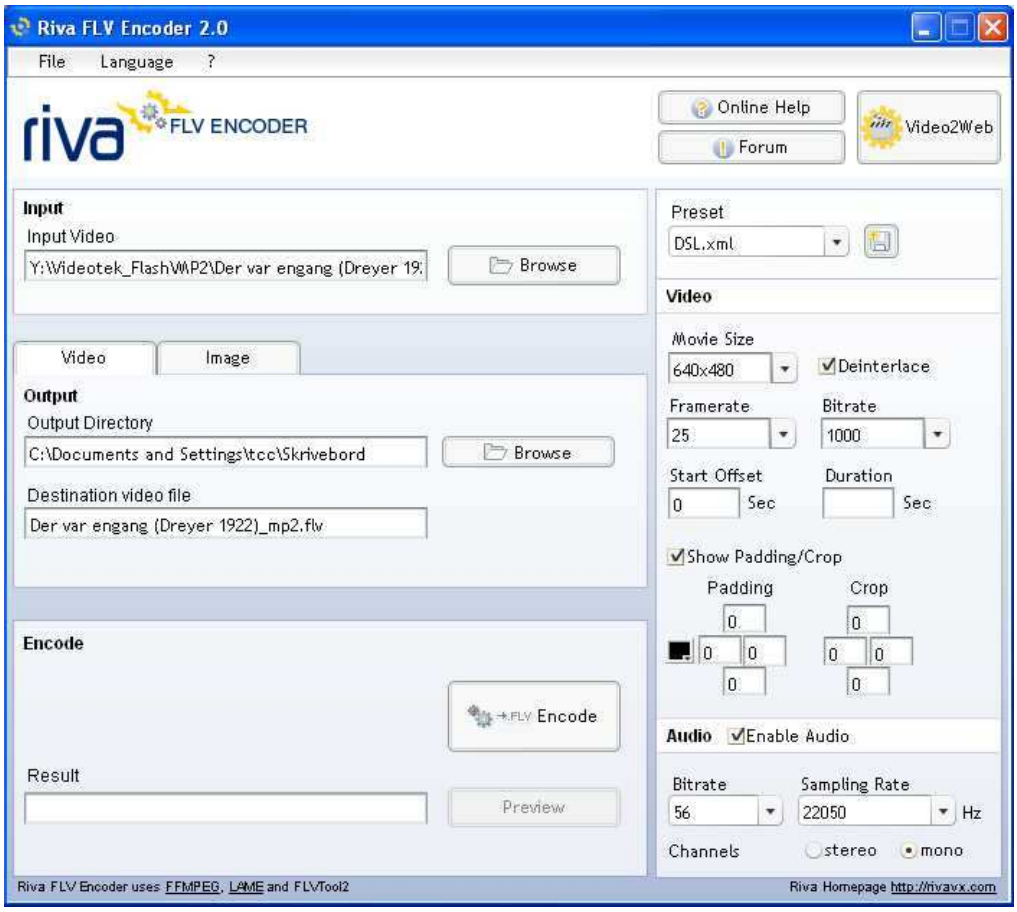
File → encoder → stream file

Encoding from file to flash

Software for encoding of flash files come in a broad range of flavours, from off the shelf full AV-suites to individual flash converters dedicated to convert from a specific file format into flash. It is possible to buy proper software such as Sorenson Squeeze, Flix Pro, or the cheaper AVS Videoconverter. However, there are also many free converters. Two of these are introduced below and will be demonstrated at the M4.3 workshop.

Riva FLV Encoder 2.0

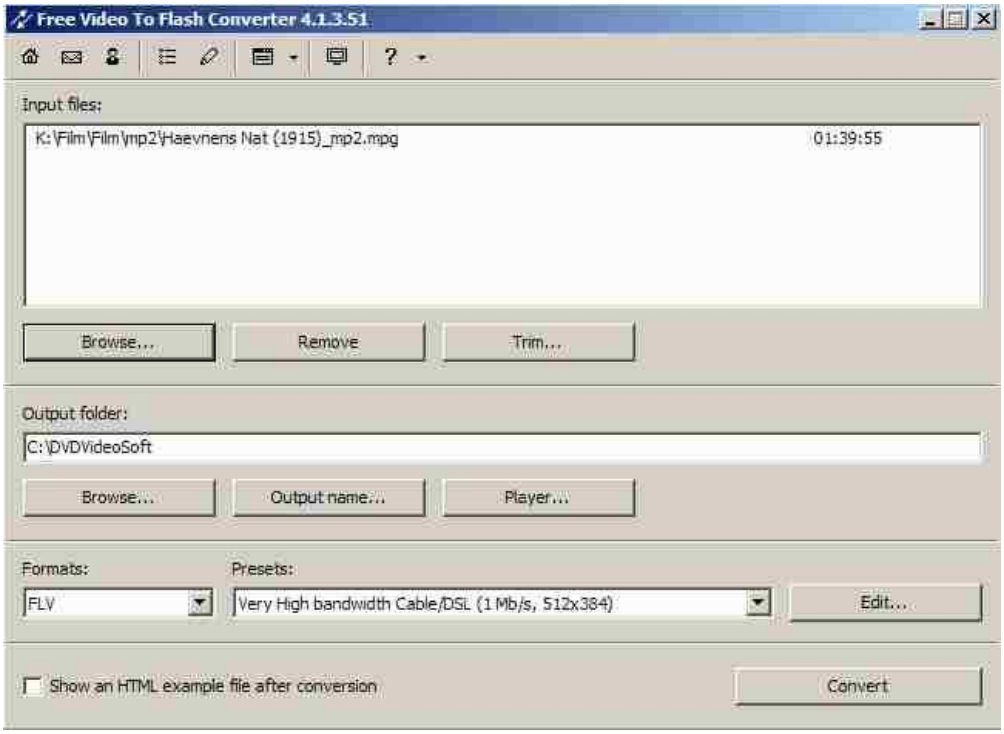
The Riva FLV Encoder 2.0 offers a number of settings within the range of EFG. The main draw-back of this encoder is the lack of batch encoding. However, as a single file encoder it is a capable and simple interface.



The Encoder is available as freeware and can be downloaded here:
http://rivavx.com/index.php?downloads0

DVDVideosoft

DVDVideosoft Flash Converter is an encoder, which delivers the possibility to do batch conversions. It both has a number of preset file settings, as well as the possibility of creating new custom file settings. It is part of a software package with several other converters. It offers several settings within the specifications suggested for the EFG.



The DVDVideosof Flash Converter is freeware and can be downloaded here:
http://www.dvdvideosoft.com/products/dvd/Free-Video-to-Flash-Converter.htm

The main objective is to find a setting, which both falls within the EFG requirements, but also makes sense in connection with the local wishes as far as image quality and system capacity is concerned. The EFG requirement is a compromise to both ensure a certain minimum quality, but also a pragmatic approach to the typical web-user, who should meet a relatively consistent level of content when using the EFG portal.

4 Distribution

The European Film Gateway infrastructure is based on a de-centralized model, where the contributing archives retain control over the access to the digitized resources. A link from the European Film gateway to the resource should provide seamless access to the resource without the user being aware of the search engine and content being hosted in different locations.

Delivery of digital objects can be done through embedded solutions that provide streaming files or through download of the digital objects. The distribution of production grade content is foreseen to be facilitated by direct contact to the providing institution or rights holder; the contact information could be part of the metadata displayed.

Streaming/embedded solutions will be provided as an integrated part of the EFG website, and will facilitate flash video and jpeg image display. Text documents will only be presented as download items. The contribution of download items of moving images, sound and images as well as documents can be contained in EFG, but will probably be directed and presented through the contributing archive's own website.

For distribution EFG will prefer the following formats and specifications:

Embedded/integrated on the EFG website, with delivery/streaming from providing institution:

4.1 Movies

The desired file format for moving images to be shown directly in the EFG interface is Flash video, which, due to its wide use, can be played independently of the systems and different browser software in the field.

Flash video (file extension .flv):

Desired minimum size: 400x300 px	Desired minimum data rate: 500kbs
Maximum size: 768x576 px	Maximum data rate: 2Mbit/s

Additionally, Windows Media Videos will also be supported. However, this is not the first choice because it very much depends on whether the Windows Media Player is installed on the user's local PC, which is not always the case.

4.2 Images

Most of the images received in EFG will already be available in JPEG format, which is also the preferred format. A desired minimum size is set up, but EFG will also be able to incorporate smaller pictures.

Jpeg (file extension .jpg):

Desired minimum size: 300 px	Preferable: 600 px

Download options in general is an attractive alternative to streaming content. Download files allow users to take the content with them on portable devices and further re-purpose the content for entertainment and educational use. However, due to IPR issues it may not always be possible to offer files for download. The listed formats are some of the most widely used and supported. As long as the file type, size and quality is visible to the user, there is virtually no limit to the size or range. This list is a short list, and many other formats may be included in the project, even if they will probably have many of the same specifications: broad acceptance and support in the most common players, sensible size and quality for current web bandwidth. In general lower quality and more obtrusive inclusion of watermarks will

prevent further use. The more restrictive the material is displayed and offered the lower the "risk" of re-purposing, but also the lower the general use and attractiveness of the content and service. When giving free access to content, one may hinder the abuse of content, or offer content freely or under a Creative Commons license or the like. Format choice will be a strategic decision made at the local participant level, but EFG will steer the offered content and formats in as uniform a direction as possible.

For download, virtually any video file format is possible. Users can convert the listed formats into their preferred formats, and other formats may be offered as their use increases, or as the partners find relevant. The concrete choice of format will be a matter of which control is expressed over the content, and which use the distribution is aiming for. There is no doubt that the higher the quality and the more open the format, the greater the use of the content will be.

mpeg4	= ipod compatible
mpeg1	= vcd compatible
mpeg2	= dvd compatible

For images, sound and text, the most likely formats are listed below. Some partners will prefer to offer documents as jpeg images or PDF files to keep authenticity, whereas others might wish to OCR scan and offer text files in order to be able to do searches in the written content. Again, the better the quality and openness of the format, the more duplication and widespread the (uncontrolled) use will be.

Images	➢ JPG ➢ (TIF)
Sound	➢ WAV ➢ MP3
Text	➢ PDF ➢ (TXT)

5 Off-site hosting of digital content on reelport

Not all EFG partner archives have their content available online on institutional web spaces. Reelport, based in Cologne, will offer support to those archives by providing a web space, where streaming videos can be uploaded and made accessible on the internet. This web space mainly has three functions:

a) to allow the upload and streaming of videos

b) to display metadata, which describes the content and technical specifications of the video as well as the provider of the video etc.
c) to allow exporting metadata and URLs of the streaming video in valid XML

There will be two approaches to upload video files to the EFG web space:

a) archives can upload their videos directly to the web platform via a user interface on www.reelport.de (preferred)
b) archives send DigiBetas to reelport where they will be converted into flash format and uploaded directly by reelport to the web platform. (exception)

The EFG space, which will be provided by reelport, will have a corporate design connected to the design of the EFG web portal. The web space can be customized for the individual archives by displaying their logo together with the streaming video and metadata.

5.1 Upload of videos

To enhance the efficiency of the upload process, the archival operator will have the possibility to upload several videos at the same time. It is suggested to allow the upload of a minimum of 10 videos simultaneously, preferrably more. When logging on to the administrative area, the administrator will be able to choose whether he wants to upload videos individually or in bulk.

In case of the bulk upload, videos can be added to a list and the archival operator can then start the upload of all videos at the same time. This will reduce the work of the archival operator considerably. Once the files have been uploaded, it will be possible to add filmographic data to the respective video.

Videos can also be uploaded individually. In this case filmographic data can be inserted while the video is being uploaded.

5.2 Formats and specifications

As mentioned before, the desired file format for videos to be shown directly in the EFG interface is flash video. Archival operators should consider the indications made in the previous chapters before uploading the videos. However, even though the internal player offered in EFG has the limitations named in chapter 4.1, reelport will be able to offer the archives to upload their video files in larger resolution, if wanted. As from reelport's side there are no restrictions regarding the size of the video file.

Moreover, reelport will offer solutions for watermarking and Digital Rights Management. As the rights situation usually varies in the various archives, the individual requests of the archive will be taken into account. Generally, the archive will have to consider in how far they want to apply these safety mechanisms, which will surely protect their content from unwanted

re-use, but at the same time impair that the usability – and therefore attractiveness – of the content to the user is affected in a negative way.

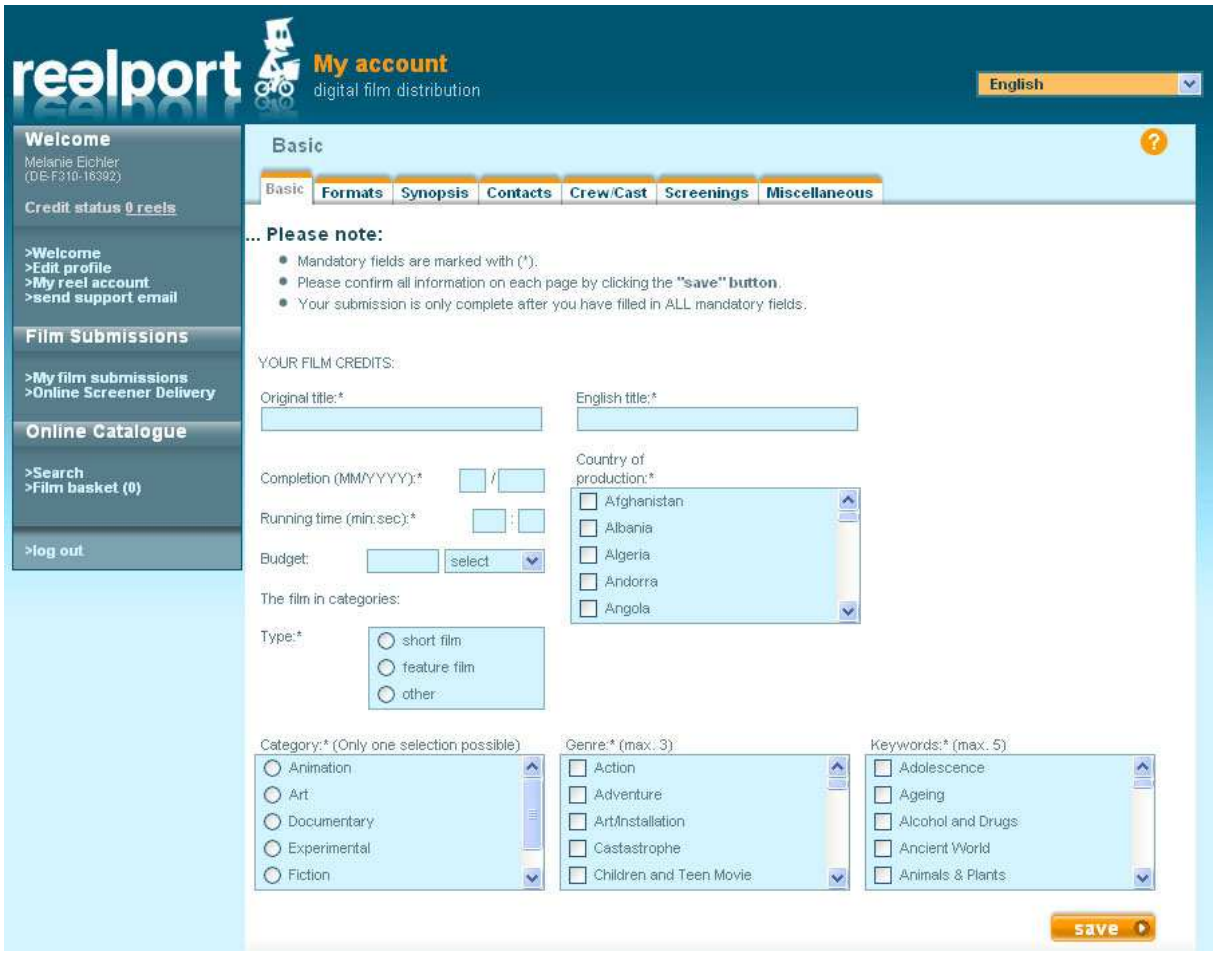
5.3 Requirements for the archival users

Each EFG content provider using the EFG space needs an individual account, which gives them access to the upload area as well as to a CMS, where filmographic data can be added, edited and deleted manually by an archival operator. The CMS shall include the following fields, in which data can be inserted:

Displayed Data	Requirements for fields
Titles	➢ free text: Multiple entries possible, each titles requires a separate field ➢ drop down menu for title type (with vocabulary from the EFG value list)
Genre	Drop down menu (with vocabulary from the EFG value list)
Country	Drop down menu (ISO 3166-1)
Year	Free text within syntactic rules
Runtime	Free text within syntactic rules
Description	Free text
Keywords	Free text
Collection	Free text
Provider	Drop down menu with names of the partner institutions
Rights	Free text
Agent	➢ Free text for name ➢ drop down menu for agent's function (with vocabulary from the EFG value list). ➢ possibility to rank the agents ➢ drop down menu for the agent's type (person, corporate body, group) ➢ For production companies a field for the domicile is also necessary
Colour	Free text within syntactic rules
Sound	Free text within syntactic rules
Language	Drop down menu (ISO 639-1)

Specifications for the values to be included in the drop-down menus will be provided by DIF. For archives, it is not mandatory to provide all information listed above, however, they are strongly advised to make as much data available as possible.

The current reelport interface will be adapted to the needs of EFG. However, the site's current functionalities will be the basis for further adjustments with regard to the data entry fields:



Screenshot: regular reelport interface for data entry

Where manual data entry seems not efficient due to a high number of film works, reelport also offers the possibility to upload XML exports. In this case, the uploaded data can be matched to the uploaded video via IDs. Entries added via this automatic procedure can be edited and enriched manually by the archival operator.

All data added – either automatically or manually – to this CMS will be exportable in valid XML, so that it then can be ingested in the EFG database. The rules for the according mapping will be written by DIF. There will be the possibility to have full exports of all data inserted as well as selective exports including data from a selected archive only.

5.4 EFG archives using the reelport web space

Six archives are likely to make use of the web space set up by reelport. These are:

- Lichtspiel Bern (LKB): ca. 160 clips
- Czech National Film Archive (NFA): ca. 1250 films (DigiBeta)
- Lithuanian Central State Archive (LCA): ca. 100 films
- Nederlands Filmmuseum (NFM): ca. 3.000 films
- Deutsches Filminstitut (DIF): ca. 100 films
- Greek Film Archive (TTE): ca. 100 films
- Portuguese Cinémathèque (CP)

The indications are of preliminary nature and not all confirmed yet. Exact numbers will be available in the near future.

6 Digital storage and strategies for long term retrieval of data

Even though digital preservation does have overlapping principles with analogue preservation, there are some major differences. One difference between analogue and digital materials is that analogue film elements are expensive to produce, but have a relatively low passive conservation cost in the long term. Digital media may be cheaper to produce, but demand a constant cost for ongoing migration and storage. Digital media furthermore have increased risk and complexity, since the content is not only subject to the carrier's decay, but also to format and software obsolescence.

Digital preservation files should be stored in a trusted digital repository. The principles behind such a storage system is to store files in two different physical locations, on two different carrier systems. The typical infrastructure is to have a harddrive master server with tape backup, as well as an additional off site tape backup. For full film quality preservation files, a tape solution might be chosen for the master files due to the lower cost of tape storage. Film preservation files would typically only need to be online for migration, since compressed formats would be used for distribution files.

Different dedicated systems would be established for preservation storage and distribution delivery, since the requirements for the uses are widely different. The preservation storage would need to be very reliable, robust, and have regular replication and safety features, but would not necessarily need to have files available online or with great speed. Users would be limited to internal access. The distribution webserver would not need the same data reliability, since the distribution files would not be unique. However, the distribution server would need to have smaller files readily available to potentially a great number of external users.

This type of setup is described in the EDCINE project. However, where the EDCINE project looked into the creation and distribution of theatrical quality masters, the EFG only deals with broadcast quality. Nevertheless, the parallel can be drawn that EFG proposes a MPEG2 or MPEG4 (6Mbit/s) as the intermediate master, from which a flash is produced for access. A master archive file can also be created. However, this needs to be of much higher

specifications (50Mbit/s+) if it is to serve as an actual preservation master of the cultural content in SD.

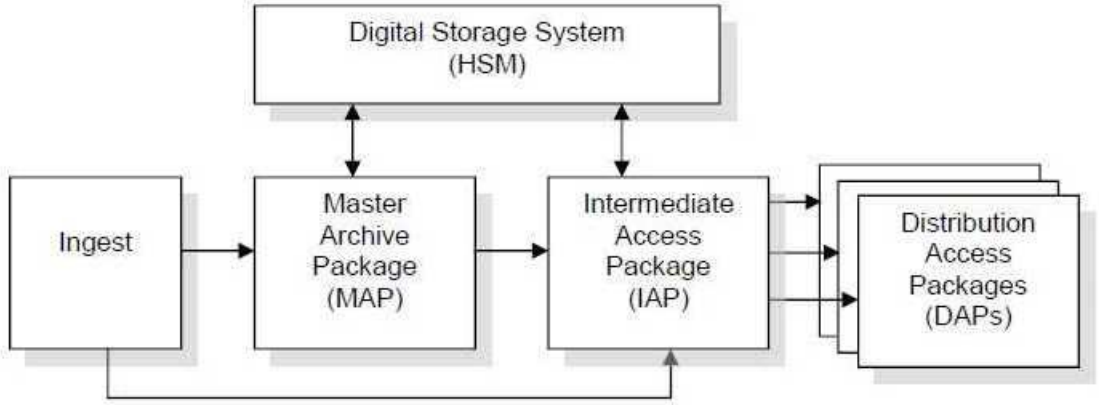


Figure from EDCine/Archives – Project Summary /v.2 October 2008. Page. 5.

It is worth keeping in mind that the total amount of data created for the EFG project is a sizeable storage responsibility:

The content listed in the DoW (p. 12) amounts to:

- About 69.000 films or film clips with a total duration of approximately 24.000 hours.
- Approximately 660.000 images including film stills, posters, photos, drawings etc.
- More than 900 items of sound material
- About 60.000 text archive documents containing approximately 300.000 pages.

The underlying **master files** comprise:

- 24.000 hours moving images (6Mbit/s): 24.000 x 3 GB = 72.000 GB = 72 TB
- 660.000 images (TIFF): 660.000 x 50 MB = 33.000 GB = 33 TB
- 900 sound items: 900 x 1 GB = 900 GB = 0,9 TB
- 300.000 pages text: 300.000 x 50 MB = 15.000 GB = 15 TB

Total master storage: 120,9 TB

(Note that actual storage requirement, including backup, will exceed this number, so the actual figure is likely to be close to three times bigger, thus creating an approximate storage need of 400 TB.)

The equivalent/parallel **web files** will comprise:

- 24.000 hours moving images (1 Mbit/s) = 24.000 x 0,5 GB = 12.000 GB = 12 TB
- 660.000 images (jpeg): 660.000 x 20 KB = 13,2 GB
- 900 sound items: 900 x 100 MB = 90 GB

- 300.000 pages text: 300.000 x 20 KB = 6 GB

Total web file storage: 12,1 TB.

These approximately 790.000 digitized items consist of more than 100 TB unique data, which at a preservation cost in a range of 500 – 5000 Euro per TB per year will cost between 50.000 and 400.000 Euro per year to maintain.

This aspect is very important to keep in mind when planning for the sustainability of the project: Due to these considerations, the federated approach has been adopted. Only thumbnails of images and videos are to be stored on the central EFG server to keep hosting and maintenance costs as low as possible.

The downside of this is, as indicated above, that a variety of formats, sizes and resolution qualities have to be respected in the EFG interface. The recommendations of this deliverable, though, are intended to serve as guidelines for existing and future EFG consortium members to produce files of recommended format, size and quality in order to enable the EFG portal to best serve its users.

7 Appendix

Approximate file sizes and qualities

Moving images:

- 4K, 4096x1714 (1:2.39), 4TB/hr, uncompressed
- 2K, 2048x1107 (1:1.85), 1TB/hr, uncompressed
- HD, 1920x1080, (16:9), 200-500 GB/hr
- SD, 720x576, (4:3), 30-60 GB/hr.
- mpeg2 (SD), DVD compatible, 768x576, 6Mbit/s, 3GB/hr, inter frame compression (mpeg2 codec)
- mpeg1, VCD compatible, 384x288, 1.15Mbit/s, 500MB/hr, inter frame compression (mpeg1 codec)
- mpeg4, H.264, MPEG4 AVC High Profile (HiP), Blu-ray compatible, 36Mbit/s, 18GB/hr
- mpeg4, H.264, MPEG4 AVC uncompressed (SD), 30-60GB/hr
- flash, 768x576, 2Mbit/s, 1GB/hr
- flash, 768x576, 1Mbit/s, 500MB/hr

Images/Photos:

Preservation: TIFF (4000x3000 pixels, 48 bit, uncompressed), approx. 50 MB

Literature:

Adelstein, Peter (2004). IPI Media Storage Quick Reference. Rochester: Image Permanence Institute.
http://www.imagepermanenceinstitute.org/shtml_sub/MSQR.pdf

EDCINE website:
http://www.cinematek.be/edcine/download/EDCine_Summary_v2.pdf

Palm, Jonas (2004). The Digital Black Hole.
http://www.tape-online.net/docs/Palm_Black_Hole.pdf

Shefter, Milt & Maltz, Andy (2007). The Digital Dilemma. Los Angeles: The Science and Technology Council of the Academy of Motion Picture Arts and Sciences.

Websites on related issues:

Digital Preservation:

TAPE Project:
http://www.tape-online.net/publications.html

PrestoSpace:
http://prestospace.org/

US Library of Congress:
http://www.digitalpreservation.gov/

Stanford University:
http://www-sul.stanford.edu/depts/dlss/services/serdigpres.htm
http://206.180.235.135/bytopic/video/

Film Preservation:

The Image Permanence Institute Preservation Calculator.
Free download: http://www.climatenotebook.org/Tfp/Software/Software_PresCalc.html

The Film Preservation Guide (2004). San Francisco: The National Film Preservation Foundation.
Free download: http://www.filmpreservation.org/

Jean-Louis Bigourdan (2002). Film Storage Studies - Recent Findings.
In: Preserve then show (2002) Danish Film Institute, Copenhagen. Editors: Dan Nissen, Lisbeth Richter Larsen, Thomas C. Christensen and Jesper Stub Johnsen.

Collection Priorisation:

NFSA Journal vol. 1, No. 1, Spring 2006.
Free download: http://www.nfsa.afc.gov.au/about_us/journal.html