DIGICULT

Digital presentation and preservation of intangible cultural heritage

O1 - Framework for common standards and models for digitization, presentation and preservation of intangible cultural heritage

DATA MODEL

Co-funded by the Erasmus+ Programme of the European Union

Erasmus+ Programme: KA2 – Strategic Partnership – VET

Ref. no. 2019-1-BG01-KA202-062231

Elaborated by	BFU (Bulgaria)
Activity related	Data model definition
Deliverable N° and title	O1 - Framework for common standards and models for digitization, presentation and preservation of intangible cultural heritage

Contents

1 Text - models and file formats 5
1.1 PDF format 5
1.2 Conclusion 6
2 Audio - models and file formats 7
2.1 Compression 7
2.2 Early formats for audio storage - WAV 8
2.3 MPEG audio: MP3 and OGG 8
2.4 MP3 file format 8
2.5 Conclusion 9
3 Image – models and file formats 10
3.1 RGB 10
3.2 CMYK 11
3.3 Advantages and disadvantages of raster graphics 11
3.4 PNG 12
3.4.1 Compression 13
3.4.2 Filtering 13
3.5 JPEG Standard 13
3.6 TIFF File Format 14
3.6.1 Tags 15
4 Video 16
4.1 Compression 16
4.2 Media codecs 16
4.3 MPEG Standard 16
4.4 MPEG-4 17
5 Metadata and standards 19
5.1 Dublin Core Standard 19
5.2 EDM Specification 20

5.3. Audio metadata 21
5.4. Multimedia metadata 21
5.4.1 MPEG-7 Standard 21
5.4.2. MPEG 21 Standard 23
6 Multimedia Libraries for Intangible Cultural Heritage 27
6.1. Physical location of multimedia objects 28
6.1.1. Texts 28
6.1.2. Images 29
6.1.3. Audio 30
6.1.4. Video 30
6.2. Physical location of files 30
7 Text repositories and metadata. Search and storage 32
7.1. Audio libraries 33
7.2. Image storage and search 33
7.3. Video data storage and search 33
7.4. Multimedia library model 34

List of Figures

Figure 1. Library for multimedia objects - basic concept 31
Figure 2. Library for multimedia objects - extended model 31
Figure 3. General scheme of a system for searching and retrieving multimedia objects from MPEG-7 based storage 34
Figure 4. Scheme of a digital object according to the MPEG-21 standard 35
Figure 5.Scheme of a first level digital object 36

1 Text - models and file formats

1.1. PDF format

Portable Document Format or PDF [ISO - PDF] is a file format used to present documents in a way that is independent of applied software, hardware and operating systems. Each PDF file encapsulates a complete description of a fixed layout, including text, fonts, graphic files, and other necessary information.

Portable document format (PDF) is actually an inter-platform electronic document format developed by Adobe Systems that uses several PostScript features. First of all, it is intended for presentation of polygraphic materials in electronic form. A significant number of modern professional printed materials have hardware support for PDF format, which allows you to print documents in this format without using software.

The PDF was officially adopted as an open standard in 2008 and was published by the International Organization of Standardization under the number ISO 32000 - 1: 2008.

The PDF format combines the following three technologies:

- A subset of the PostScript page description programming language that generates views and graphics.
- Font embedding / replacement system that allows used fonts to be completed with documents.
- System for structured storage of all elements in a single file, with application of data compression techniques where necessary.

The PDF format allows you to implement the necessary fonts, vector and raster images, forms and multimedia inserts. It supports RGB, CMYK, Grayscale, Lab, Duotone, Bitmap, several types of raster image compression. It has its own technical printing formats: PDF / X-1a, PDF / X-3. It includes a mechanism for electronic signatures for protection and authentication of documents. A large number of accompanying documents are distributed in this format.

PDF files are divided into four parts: header, objects, cross-references and trailer. The header is the first part. The objects section follows and is the largest part of the file. Then the cross-references section follows, giving the position of each object in the file as a byte offset from the beginning. The PDF trailer specifies how the application that reads the PDF document should find the cross-references comparison table and other special objects. All PDF readers should start reading the PDF file from the end of the file.

1.2. Conclusion

Textual information is a key component of multimedia libraries, and this determines the crucial importance it has in terms of their creation and maintenance. As humanity has so far gathered a huge array of textual information, the question arises how to digitize and present it in order to ensure regulated access in a global aspect.

Plain text is a text format that contains only the text - its semantics, avoiding any other additional and often redundant in terms of working with the text information. For this reason, the processing of textual content can be simplified and automated as the layout and typography do not affect the process itself. On the other hand, the use of appropriate typography and formatting is essential for users. Relevant visualization of textual information allows people to perceive it more easily and assimilate it faster.

Providing textual information to end users through the use of plain text is inappropriate, as this format has very limited possibilities in terms of providing appropriate visualization of the texts. Formats such as XML, HTML and PDF are also among the potential solutions to the problem of visualizing textual information.

2 Audio - models and file formats

2.1. Compression

Compression techniques have been developed to further process information content in order to reduce file size without losing important information. From the point of view of providing content from digital libraries in a networked environment, file size is paramount. Naturally, smaller files are provided faster. When choosing a method for compressing files, it is important to take into account not only the achieved level of file volume reduction, but also the time and resources required to perform the processes of compression and decompression.

Compression can be performed without any loss of information, which means that the decompressed result is exactly the same file as the original. In most cases, compression leads to losses in terms of file contents, i.e. in these cases there is no guarantee that the decompressed file is the same as the original. An important condition in the application of compression schemes with losses is that these losses are small and insignificant enough and, if possible, not noticeable. Compression techniques leading to losses achieve significant results in terms of reducing file size, but there is also a slight deterioration in quality.

It is important to note, however, that when it comes to text compression, only the application of lossless compression schemes and techniques makes sense. With regard to digital images and audio information, the application of lossless compression is associated with long and resource-intensive processing, while having a negligible effect. When applying lossy compression techniques, the time and resources required are less, and the losses themselves in terms of content are negligible and inconspicuous.

In today's world of Internet and global information exchange, compression techniques with loss of information are mainly used. However, when small, insensitive to human receptors distortions are resolved, lossy compression methods can achieve impressive results. At the physical level, signals include noise that is essentially uncompressed. The paradox is that lossy compression sometimes not only reduces file size, but in some cases actually improves the signal by reducing noise. When applying lossy compression techniques such as jpg and mp3, remarkable results can be achieved for digital images and audio - initial file size is reduced many times over.

However, accurate representations (excluding losses) are essential in some cases. For example, in research units where analyzes are performed on the basis of registered defects in the materials, research hypotheses can be formulated and / or important conclusions can be drawn.

2.2. Early formats for audio storage - WAV

The WAVE format, or more commonly known as WAV (Waveform Audio Format) due to file name extension, is the standard format for Microsoft and IBM audio files for audio storage. This is an application of Resource Interchange File Format (RIFF) bitstream format method for storing data in containers ("chunk").

Although a WAV file may contain compressed audio, the most common WAV audio format is uncompressed LPCM (Linear Impulse Code Modulation)

Because LPCM is uncompressed and retains everything from audio, professional users or audio experts can use the WAV format with LPCM for maximum quality. WAV files can be edited and manipulated relatively easily using software.

2.3. MPEG audio: MP3 and OGG

MP3 (formally MPEG-1 audio layer III or MPEG-2 audio layer III) is an audio format for digital audio. MP3 was designed by the Moving Picture Experts Group (MPEG) as part of the MPEG-1 and later MPEG-2 standards.

Originally defined as the third audio format of the MPEG-1 standard, it has been preserved and further developed - additional bit rate and support for multiple audio channels are defined as the third audio format of the next MPEG-2 standard.

2.4. MP3 file format

MP3 (or mp3) as a file format usually refers to files containing a simple stream of MPEG-1 audio and video encoded data, without other additions to the MP3 standard. MP3 was designed by the Moving Picture Experts Group (MPEG) as part of the MPEG-1 and later MPEG-2 standards.

In the aspects of MP3 related to audio compression, MP3 uses data loss for compression. Inaccurate approximations and partial data erasure are used to encrypt the data. This allows a significant reduction in file size compared to uncompressed audio. The combination of small size and acceptable quality led to a boom in music distribution. With the advent of portable media players, including smart phones, MP3 is almost universal.

MP3 compression works by reducing (or zooming in) the accuracy of some components of sound that are considered beyond the hearing ability of most people. This phenomenon is known as acoustic masking: the human auditory system cannot perceive low-amplitude components that are dominated by neighboring high-amplitude components. Temporal masking is related to the following phenomenon: listeners do not hear faint sounds that occur temporarily next to louder ones. Again, these elements can be encoded with lower accuracy or discarded entirely by the encoded signal.

Compared to CD-quality digital recording, MP3 compression can usually reduce the size by 75 to 95%. For example, MP3 encoded at a constant bit rate of 128 kbit / s would result in a file with approximately 9% of the size of the original CD audio.

Although MPEG is an open standard, various companies hold patents for some of the techniques it uses.

2.5. Conclusion

Taking into account the characteristics and capabilities of the considered formats, we can say that the MP3 format is suitable for use when performing a preview of a record. On the other hand, we must emphasize that this format is not the best choice when it comes to archiving and storing audio information in digital libraries. The situation is similar, in which the main goal is the detailed study of sound. In this case, MP3 is also not particularly suitable. For example, the human ear perceives sounds in a limited frequency range, but hearing in some animals is much more developed. If the aim of the work is the detailed analysis of the sound, then a format without loss of quality should be used.

It is recommended to use formats without loss of quality for storage of audio information in digital media repositories. To provide opportunities for quick listening and searching, each such file should be associated with another that keeps the audio information in compressed form. Our suggestion is to use the WAV format for long-term storage and protection of audio information and its associated OGG file for easier searching and listening. As the price of information carriers is constantly decreasing and is now quite affordable, so the cost of storing a unit of information volume has also become cheaper in recent years, such a solution is easily feasible and cost-effective.

3 Image – models and file formats

Raster images are visualized by a raster, i.e. a rectangular grid or matrix of elements called pixels. Each pixel has its own value that corresponds to its color. The dimensions of the matrix determine the resolution of the image. It is measured by the notation dpi (dots per inch - number of dots per inch distance). The other way to indicate a resolution is by presenting the total number of dots contained horizontally and vertically - for example 800x600, etc. The standard resolution of the monitors is 96 dpi. However, if it is a material that will be printed, then the standard is a minimum of 300 dpi.

There are different methods for setting (describing) the color of a pixel. These methods are called color schemes.

Some of the most common color schemes are as follows:

- Black and white - black and white color scheme - only black and white - occupied memory - 1 bit.
- Grayscale - halftone color scheme - grayscale - memory consumption: 1 byte.
- Color - color scheme for full color presentation.

There are several basic ways to represent colors. The following sections describe the most popular ones.

3.1. RGB

RGB is short for Red, Green and Blue - the three components that define a color.

At zero intensity of the three colors, black is obtained, and at maximum - white. Monitors use this representation. It is used to represent electronic devices.

Additive color mixing - occurs when working with color streams of light and is based on the three main colors: red, green and blue. When mixing them, light color tones are obtained, i.e. with a combination of red and green, yellow is obtained, of green and blue - cyan (blue-green), and of blue and red - purple.

3.2. CMYK

Subtractive color mixing is typical when working with color substances (for printing, for example). These substances absorb certain wavelengths and at the same time reflect other wavelengths. CMYK is a color model (abstract color quantification model) used in modern color printing, including screen printing, offset printing, and PC printing. The abbreviation is formed by letters of the English words Cyan (cyan - mainly blue), Magenta (purple-red color, tending to red), Yellow (yellow), Key or blacK (for black).

The numerical value of each color element of CMYK system is between 0% and 100%, where at 0% the specified color element is not reproduced, and at 100% it is reproduced completely. The percentage value indicates the saturation of the given color element. It uses 3 basic colors in different concentrations - blue-green (Cyan), purple (Magenta) and yellow (Yellow). The zero concentration of the three primary colors gives white, and their maximum concentration – black color.

3.3. Advantages and disadvantages of raster graphics

If we have A4 bitmap format raster graphics with a resolution of 300 dpi and a 24-bit full color image, then the file size will be 30 MB - calculated as the product of the total number of dots of the image by 24. At even higher resolution, the file size increases even more. From this it immediately becomes clear that the storage of such information is extremely resource-intensive. Of course, as with audio storage, we can store images with loss of information and without loss of information.

Scanned images can contain digital noise called color, as well as distortions caused by the digitization process itself. A lossy compression scheme that can eliminate noise, achieves file minimization while increasing image quality. Such techniques can be used to obtain extremely compact images, especially for gray scale and photographic images. Depending on the content of the image, each pixel can be compressed from its original with a representation, in which one to four bytes are reduced to only one bit and with a barely noticeable loss of quality.

However, there are situations in which accurate performance is essential and compression should be lossless, even without clearing color noise in any case.

One basic approach to image compression is to apply a standard program designed for general compression of computer files, such as zip. They treat the file as a linear sequence of bytes.

Compression can be significantly improved by recognizing the two-dimensional structure of the image and by using the spatial characteristics of the data. For example, most images contain both large spaces whose color and texture are constant or slowly changing, and regions where abrupt changes such as lines or edges are observed. Sophisticated algorithms identify these types of regions and use the information from the analysis to perform image compression.

In what follows, we will look at the PNG format for compression without loss of quality, comparing some of its features with those of the JPEG format, which was developed for lossy information (quality) compression for photographic images. Both formats are supported by browsers and related applications.

The TIFF format, which is often used in digital libraries as a working format during digitization itself and for archiving information, will also be considered. The format allows both storing images with loss of information and storing an image without loss.

The last chapters of the section discuss the scalable vector graphics (SVG) format, which allows images to be defined in a structured way and provides a richer set of options for developing search and browsing operations in a digital library, than purely raster formats.

3.4. PNG

Portable Network Graphics (PNG) is a raster graphics format that supports lossless data compression. PNG was created as an improved, non-proprietary substitute of Graphic Interchange Format (GIF) and is among the most widely used image compression formats on the Internet.

PNG supports palette-based images (24-bit palettes for RGB or 32-bit palettes for RGBA colors), grayscale images (with or without an alpha channel for transparency), and full-color RGB / RGBA images with or without an alpha channel.

PNG is designed to visualize images on the Internet, not professional-grade graphics, and therefore does not support color models such as CMYK. The PNG file contains a single image in an extensible structure of "chunks" encoding key pixels and other information, such as text comments and CRCs documented in RFC 2083. The files have the file extension PNG or png and MIME type image / png.

PNG was published in RFC 2083 in 1997 and as an ISO / IEC standard in 2004

Characteristic of PNG is that it provides better compression as it works by recognizing the two-dimensional structure of the image and defines filters that can be applied to the pixel values before compression. The horizontal difference filter takes the previous pixel value from the current one, so the differences in the pixels are actually encoded, not the pixel values.

Compared to GIF, PNG improves compression by 10 to 30%, depending on the specific image being encoded.

- Image transparency

PNG offers a variety of transparency options. For true color or gray scale images, they can be declared transparent or an alpha channel can be added (allowing the use of a percentage of transparency). For palette images, alpha values can be added to the entries in the palette.

3.4.1. Compression

PNG uses a two-step compression process:

- pre-compression: filtering (forecasting);
- compression DEFLATE.

PNG uses DEFLATE, a non-proprietary data compression algorithm that uses a combination of LZ77 and Huffman encoding.

3.4.2. Filtering

Before applying DEFLATE, the data is converted using a "prediction" method: for the image as a whole, one filter method is used, while for each row of the image, a filter type is selected to transform the data to become more compressible.

There is only one filtering method in the current PNG specification (denoted as method 0).

3.5. JPEG Standard

JPEG (Joint Photographic Experts Group) is a comprehensive standard for presenting continuous-tone compressed images. It was approved in 1994. It is a general purpose standard and is the basis of many graphic applications and image-related services. Since its inception, the format has become the standard for photographic images. The JPEG compression algorithm works well enough to provide excellent image quality at about 1 bit / pixel.

JPEG is a format that can be used at 64 kbit / s speed, making it suitable for interactive use. Many graphics applications offer image quality adjustment in the range of 1 to 12.

JPEG compression is used in many image files' formats. JPEG is the most common image format used by digital cameras and other photographic imaging devices. This format, together with the JPEG / JFIF format, are the most common formats for storing and transmitting photographic images on the Internet. Maximum image size is of 65,535 × 65,535 pixels or up to 4 giga pixels at a 1: 1 aspect ratio.

JPEG compression involves removing 75% of the image color information. Thus, the storage data is reduced by half, and at the same time the human eye hardly detects a difference with the original.

JPEG is not suitable for encoding drawings or text where strong contrasts can cause noticeable distortions. For such images it is better to use TIFF or PNG.

Because quality is lost with each JPEG compression, the compression method is not suitable for materials that require 100% fidelity, as well as for materials that will be repeatedly compressed.

3.6. TIFF File Format

Tagged Image File Format, or TIFF or TIF, is a file format for storing bitmap graphics, especially popular in the art world. The format was created by Aldus Corporation with the copyright of Adobe Systems, acquired after the purchase of Aldus.

TIFF is a flexible and adaptable file format for processing images and data in a single file, including header tags (size, definition, image, data, image compression) that define the geometry of the image. The TIFF file, for example, can be a container with JPEG (lossy) and PackBits (lossless) compressed images. The TIFF file can also include vector-based drawing (outlines, crop, frames). The ability to store image data without loss of information makes TIFF useful for archiving, because unlike standard JPEG files, TIFF files use compression without loss of information and can be edited and re-saved without losing the quality of the image. This is not the case when TIFF is used as a container containing compressed JPEG.

TIFF offers the option of using LZW compression without data loss - a compression technique for reducing file size.

The TIFF file can contain any number of images - zero and more. Each image is considered a separate subfile and has an IFD describing the data. Each subfile can be saved as a separate TIFF file or can be stored with other sub-files in one TIFF file. Each bitmap and IFD subfile can be located anywhere in the TIFF file after the header, but there can only be one IFD per image.

3.6.1. Tags

You can think of a tag as a data field in the header of the file. However, while the header data field may contain only fixed-size data and is usually only at a fixed position in the header of a file, the tag may contain or point to data that is at different number of bytes and is found everywhere in IFD.

4 Video

A codec is hardware or software for encoding or decoding a data stream. The hardware codec can convert analog to digital signal and vice versa. Software one - encodes and decodes data streams.

4.1. Compression

In addition to encoding, the codec also performs compression. We can distinguish the following classes: codecs, compressing without data loss, and those compressing with loss of information.

Lossless codecs are usually used for archiving because they store all information without loss. In the same way as with the JPEG standard, if we apply multiple consecutive encoding decoding or more than one codec or coding scheme with loss of information, the quality can deteriorate significantly. The continuous reduction of media prices allows videos encoded without loss of quality to be stored at a relatively good price.

The most popular are codecs that have a loss of quality because they have high compression and small files. Again, methods of human psychological perception are used to achieve compression that is difficult to distinguish from the uncompressed original, but with a significant reduction in file size.

4.2. Media codecs

Pulse code modulation and delta modulation are commonly used in codecs. Codecs are designed to highlight some aspects of the media that need to be encoded. Too often, media data streams contain audio and video, as well as some metadata that allows audio and video to be synchronized. Each of these three streams can be managed by different programs, processes, or hardware, but in order for streaming media data to be useful in stored or transmitted form, it must be encapsulated together in container format.

4.3. MPEG Standard

Moving Picture Experts Group (MPEG) is a working group created by ISO and IEC, which set standards for audio and video compression and transmission.

MPEG standards consist of different parts. Each part covers a certain aspect of the whole specification and each standard specifies a profile and level. Profiles are designed to define a set of available tools, and the levels determine the range of appropriate values for the associated properties. MPEG standardizes the following compression formats and additional standards:

- MPEG-1: Motion picture encoding and associated audio for digital media up to about 1.5 Mbit / s (ISO / IEC 11172). It includes MPEG-1 Audio Layer III (MP3) audio layer compression format;
- – MPEG-2: Generic encoding of movies and associated audio information (ISO / IEC 13818). Transport, video and audio standards for high quality television;
- – MPEG-4: Encoding audiovisual objects. (ISO / IEC 14496) MPEG-4 uses additional coding tools with additional complexity to achieve higher compression ratios than MPEG-2.

4.4. MPEG-4

MPEG-4 is a method for determining the compression of audio and visual digital data. It was introduced at the end of 1998 and set a standard for a group of audio and video encoding formats and related technology developed by the ISO / IEC Moving Picture Experts Group (MPEG) (ISO / IEC JTC1 / SC29 / WG11) ISO / IEC 14496 - Encoding of audiovisual objects. The use of MPEG-4 includes compression of AV data for web (streaming media) and CD distribution, voice (telephone, video intercoms) and television applications.

MPEG-4 provides a series of technologies for developers, different service providers and end users:

- MPEG-4 allows various software and hardware developers to create multimedia objects with better adaptability and flexibility to improve the quality of such services and technologies as digital TV, animation, World Wide Web and their extensions.
- Data network providers can use MPEG-4 for data transparency. Using standard procedures, MPEG-4 data can be interpreted and transformed into other types of signals compatible with any available network.
- The MPEG-4 format provides the ability to interact with various animated objects.
- Standardized signaling for digital rights management, management and protection of intellectual property (IPMP).
- The MPEG-4 format can perform various functions, including the following:
- Multiplexing and synchronizing data related to media objects in such a way that they can be efficiently transported additionally via network channels.
- Interaction with the audio-visual scene which is formed by the user.

MPEG-4 provides a more comprehensive multimedia concept and is designed for use over low-bandwidth web and mobile networks. Video and audio are complemented by still images, generated graphics, sound and text. These primitive data types can be hierarchically composed into objects. Objects replace flows as a key structure, offering more expressive power than before. MPEG-4 is a huge standard. There are 33 parts! The first five again correspond to those in MPEG-1. MPEG-4 can provide a quality service that adapts to the available internet bandwidth. It allows users to interact with the scene and to implement strategies for identifying and managing intellectual property.

5 Metadata and standards

Metadata provides a common set of labels that provide information about resources, regardless of who authored them, what tools were used to create them, or where they were stored. The set of labels describes the intended way in which the resource should be used. The use of metadata makes it possible to describe, summarize and search for information resources, which in turn is necessary for the reuse of the latter.

Different societies of professionals have developed their own standards and standardized metadata dictionaries that meet their specific needs, which helps to carry out joint activities within a subject area, but often leads to incompatibilities between separate and inhomogeneous metadata schemes for different areas.

Metadata, often described as data for the data, is crucial to all forms of organized digital content. It is the means by which organization is achieved in the multimedia library. Metadata is usually considered to be structured information for a specific information resource. Information is "structured" if it can be manipulated essentially without understanding its content.

Important questions in this aspect are whether metadata is retrieved automatically from digital objects or it is done manually, what is its impact on visualization, browsing, search and maintenance, updating, sharing and upgrading. The following is a brief overview of some metadata and description standards that are of particular interest to us.

5.1. Dublin Core Standard

Dublin Core (DC) BSS - ISO 15836: 2010 is an improved digital cataloging system that makes search engines much more accurate and efficient. The Dublin Core scheme has many terms to describe resources such as web pages and media such as video and images. There is also data for physical objects such as CDs, books, works of art. The main purpose of this system is to create a powerful and convenient catalog, including various objects. It can be used for better search engine optimization. The resulting metadata can be used to quickly describe resources and to combine metadata from different standards.

The Dublin Core Metadata Initiative is an open organization committed to developing compatible online metadata standards that support a wide range of goals and business models.

There are two types of Dublin Core: Simple Dublin Core and Qualified Dublin Core. Simple Dublin Core is for simple attribute value pairs and uses 15 classic elements, while Qualified Dublin Core uses more elements to better define the data.

5.2. EDM Specification

Europeana is a European web portal that brings together digitized sites that reflect different aspects of European culture. EUROPEANA data model (hereinafter referred to as EDM for short) is as follows:

EDM Definition [EDM] - the official specification of Europeana Data Model and lists the classes and properties that can be used in Europeana. However, not all of them are currently in use.

EDM PRIMER- shows how classes and properties can be used together to model data and maintain Europeana functionality.

MAPPING GUIDELINES EDM - provide guidance for content providers who want to use EDM to present their content. They show which property is associated with which class, property definitions, data types that can be used as values, and so on. It also contains an example of original data that is converted to EDM and diagrams showing the distribution of properties among classes. The full set of classes and properties of EDM are implemented gradually and MAPPING GUIDELINES is the reference document showing the currently available ones.

EDM OBJECTS TEMPLATES - a simple wiki list showing which properties are applied to which class and in which the data types and the range of values are specified. However, these templates should be considered as working options and may not be in conformity with MAPPING GUIDELINES.

XML schema - this is the XML schema for the current EDM implementation.

EDM aims to be integration medium for collecting, linking and enriching the descriptions provided by EUROPEANA content providers. As such, it can be said to include any element (i.e. class or property) found in the content provider's description. Accounting for all these elements is clearly impossible, as they are an open set, i.e. a set that can be expanded when new providers join Europeana's information space. However, there is a well-identified set of elements that EDM uses to accomplish its task. These elements can be divided into two main categories:

1 Items reused by other standards;

2 The elements entered via EDM.

EDM uses elements from the following initiatives:

- Resource Description Framework (RDF) and the RDF Schema (RDFS) http://www.w3.org/2000/01/rdf-schema#/
- OAI Object Reuse and Exchange (ORE) (http://www.openarchives.org/ore/terms/)
- Simple Knowledge Organization System (SKOS) (http://www.w3.org/2004/02/skos/core)
- Dublin Core namespaces for properties from the elements, terms and types (http://purl.org/dc/elements/1.1 , http://purl.org/dc/terms , http://purl.org/dc/dcmitype/)
- W3C Data Catalog Vocabulary (DCAT (http://www.w3.org/tr/vocabdcat/)
- Creative Commons (CC) http://creativecommons.org/ns
- SIOC Services Ontology Module http://rdfs.org/sioc/services#3

5.3. Audio metadata

Most often, audio resources are described using external metadata, such as Dublin Core.

The MP3 format contains its own tags. The first set is known as ID3 (for "IDentify an MP3"). ID3 tags were placed at the end of the audio and were limited to 128 bytes: 3 for the initial series "TAG" which identifies the metadata, a block of 30 bytes for the title, 30 for the artist, 30 for the album, 4 for the year, 1 for the number of the song, 1 for genre and 30 for comments. The current ID3v2 tag format allows many variable-size data containers (up to 16 MB each) and can contain text, images, and technical metadata for audio data. The metadata is located at the beginning of the file. There are 39 predefined text containers, such as composer, length, copyright, publisher and title. Others provide the ability for URLs, images, karaoke texts, and more.

5.4. Multimedia metadata

5.4.1. MPEG-7 Standard

The lack of standards for multimedia descriptions is the main prerequisite for the development of the MPEG-7 standard (MPEG-7 Documentation, 2004), which offers a framework for standardized multimedia descriptions of information content. MPEG-7 /Multimedia Interface for Content Description/ is a standard for describing multimedia content, maintaining a certain degree of interpretation of the meaning of data / information. The standard offers an understandable set of audio-visual descriptive tools, metadata, structure and links defined by the standard in the form of descriptors and descriptive schemes. Descriptions should form the basis of applications, enabling efficient and effective access (search, filtering and browsing) to multimedia content (audio-visual data). The descriptors and descriptive schemes of MPEG-7 provide a possibility for standardized description of: static images, graphics, 3D models, audio, speech, video and combinations of these elements combined in a multimedia presentation [Kosch et al. 2003].

MPEG-7 takes into account key aspects included in leading standards for multimedia descriptions such as TV Anytime, Dublin Core, SMPTE Metadata Dictionary, and EBU P / Meta. While these standards target more specific applications and domains, MPEG-7 has been developed as the most general standard possible, not targeted at a specific application or area, but rather intended to provide maintenance for the widest possible range of applications.

MPEG-7 has an extremely wide range - metadata can be used to search by similarities for music, images, textures, and even to search for relationships between objects and scenarios.

MPEG-7 is based on four components: *Descriptors, Description Schemes, Description Definition Language, and Systems Tools.*

The descriptors represent the main attributes of low-level audiovisual content. They define the syntax and semantics of each characteristic (through a metadata element) and the descriptive schemes that define the structure and semantics of the relationships between the elements, which can be both descriptors and descriptive schemes. Descriptors usually represent quantitative measures linking specific features to a set of values.

Schemas define the types of descriptors that can be used and the relationships between them or between other description schemes. Descriptive schemes can be considered as models of objects and environments from the real world. In itself, a model defining the specific descriptors to be used in the description, the relationships between the descriptors or between the description schemes.

Description Definition Language (DDL) allows users to extend the predefined capabilities of descriptors and description schemas - defines the syntactic rules for defining, expressing, and combining descriptive schemas and descriptors. DDL uses XML syntax and is a form of XML Schema.

System tools are a set of basic tools for synchronizing metadata descriptions with encoding and content transmission. They maintain double-encoded images for efficient storage and transmission, a transfer mechanism (for both text and binary format), description multiplexing, synchronization of descriptions with content, management and protection of intellectual property in MPEG-7 descriptions, and more.

There are various descriptors for audio, visual and multimedia data. At the lowest level are the instantaneous values of wave strength and power, characteristics of frequency spectra, etc. At a higher level, sound effects, descriptors of melodies are described. In the visual domain, the main characteristics include color, texture, shapes for areas and contour shapes, the movement of the camera and the subject. Also characteristic is the concept of localization in time and space and these quantities can be combined in space-time trajectory, time series of video frames, etc.

Multimedia functions include low-level audiovisual attributes such as color, texture, motion, compression information, storage media, and more. An audiovisual object can be navigated either hierarchically or sequentially. For hierarchical navigation it is organized in successive levels which describe it at different levels of detail, from general to detailed. For sequential navigation, images or video frames can be arranged sequentially and possibly synchronized with audio and text so as to compile audiovisual summary information.

An important feature is that MPEG-7 metadata can be entered manually or retrieved automatically. Some functions (color, texture) can be retrieved automatically, while others (descriptive) cannot be retrieved automatically.

5.4.2. MPEG 21 Standard

The implementation of an open multimedia infrastructure requires an integrated and appropriate digital rights management system. MPEG-21 aims to ensure interoperability by focusing on how the infrastructure elements of multimedia applications connect, integrate and interact, and where there are no open standards for elements, MPEG 21 creates such.

The main premise of MPEG-21 is that each user is potentially an element of a network that includes a huge number of content providers, users and merchants, and more. In addition to client-server and peer-to-peer applications, user role flexibility is a key part of MPEG-21. Interoperability is the driving force behind all multimedia standards and is a necessary requirement for any application that requires guaranteed communication between two or more parties.

Basic concepts in MPEG-21: A digital element is a structured digital object with standard representation, identification, and metadata within the MPEG-21 framework. The user interacts in an MPEG-21 environment or uses a digital object. User roles include creators, users, rights holders, content providers, distributors, and so on. All the parties that interact within MPEG-21 are categorized as users. They assume specific rights and responsibilities according to their interaction with other users. A digital object is a combination of resources, metadata, and structure. Metadata describes data for or relating to the digital object as a whole or also to its individual resources and to the relations between the individual parts of the digital object - both resources and metadata. The digital object is the basic unit for distribution and transaction within MPEG-21. The parts of MPEG-21 are organized into several separate chapters.

1. Vision, technology and strategy: describes the multimedia framework and its architectural elements with the functional requirements for their specification.

2. Digital Object Declaration (DID) - provides a digital object declaration scheme.

3. Digital object identification (DII): as properties, type, etc.

4. Intellectual Property Management and Protection (IPMP): provides the means to reliably manage and protect network content and devices.

5. Language for the presentation of rights (REL): indicates a language that can declare the rights and access using the terms defined in the Data Rights Dictionary.

6. Data Rights Glossary (RDD): specifies the glossary of key terms needed to describe users' rights.

7. Digital Object Adaptation (DIA): identifies tools to describe the usage environment and content format characteristics that could affect access to multimedia content.

8. Reference software: includes software that executes the instruments specified in the other parts of MPEG-21.

9. File format: specifies the file format for storing and distributing digital objects.

10. Digital Object Processing (DIP): defines mechanisms for standardized and interoperable processing of information in digital objects.

11. Methods for assessment of sustainable associative technologies: document best practices in the assessment of persistent association technologies, using a common methodology. These technologies link information that identifies and describes the content directly to the material itself.

12. Resource Delivery Test in MPEG-21: provides a software-based test layer for delivering scalable media and testing / evaluating this scale.

Vision, technologies and strategy

The MPEG-21 technical report describes the multimedia framework and its architectural elements with the functional requirements for their specifications.

Declaration of the digital object

It indicates a unified and flexible abstraction and an interoperable scheme for declaring the structure and type of digital objects. Through the Declaration Language (DIDL), we can declare a digital object by specifying resources, metadata, and their relationships.

Rights Expression Language

REL is a language to express rights pertaining to digital objects, components, fragments and containers. It provides mechanisms that support the use of digital resources in publishing, distributing and using digital objects (films, music, e-books, software and others in digital form) in a way that protects digital content and preserves the rights set for specific digital content. REL also provides a flexible, interoperable mechanism to ensure that personal data is used lawfully.

Data Rights Glossary

The glossary includes a set of clear, consistent, structured, integrated and identified conditions for REL support. Defines the structure and basic terms and specifies how the terms can be defined and registered in the future. The use of the glossary facilitates the accuracy of the exchange and processing of information between the parties concerned, involved in the administration of rights and the use of digital objects. It also maintains the transformation of metadata from the terminology of one domain to another in an automated or partially automated way, with a minimum of ambiguity or loss of semantic integrity.

Digital Item Adaptation

The goal is to achieve interoperable transparent access to distributed multimedia content by shielding users from network and terminal installation and management. Not only the description of the content is essential, but also the description of its format and the environment of use, so that the content can be adapted and the user can be provided with the best content found as a result of the search.

Digital Item Processing

DID (Digital Item Declaration) is static - only the information (structure, resources and metadata) in the object is declared. There is no processing of this information. It includes only such processing as downloading a digital item, which also requires rights to do so. When receiving a digital item, users have a list of available processes (methods) that can be applied to it.

6 Multimedia Libraries for Intangible Cultural Heritage

Intangible cultural heritage is an extremely important aspect of our live. We are so used to it that we do not even notice it clearly, but take it for granted. When we go to another country or city, we mostly visit museums or galleries, but we very rarely notice that we are mainly acquainted with the intangible cultural heritage of that country. This is typical not only for non-specialists, but also for the people involved in this domain - museum workers, gallery workers and owners, etc. Most developments for multimedia libraries, for example, emphasize the digitization and search for archaeological sites, icons, paintings and other tangible valuables. This is because the very process of digitization of objects related to intangible cultural heritage is complex and versatile. Here we will consider, without claiming to be exhaustive, the methods for creating a multimedia library for intangible heritage and will offer architecture for its implementation.

Let's look at an example of intangible heritage. Such an example is the Nestinarstvo (Fire Dancing) custom, which is practiced in several villages in southeastern Bulgaria and northern Greece. The custom is registered into the list of the world's intangible cultural heritage. We speak about this custom without thinking about what it actually contains and if we have to describe it, preserve and present, what we should do and how. The custom itself contains many aspects, which is why we have chosen it as a suitable example. There are many written sources about it. There are also other descriptions of the ritual itself already at the end of the 19th and the beginning of the 20th century, which are interesting from an ethnographic point of view. The text itself is interesting to us as information, not as layout. However, layout is important when reading. Well-formed text is easier to read and perceive faster. As we know, serif fonts, which originated already in the 19th century, are read 20-30% faster than ordinary non serif fonts. For their part, however, they strain the eyes more when viewed on a monitor. Most of the texts are also on paper and have not yet been digitized. Namely, when digitizing these materials, we should use texts that are in digital form, but also to use techniques for entering information from paper and funds of various cultural institutions.

In addition to texts, many images can be added to the custom: From photographic material from the first studies, through photos of the lands, of used icons, schemas of the location of the event, etc. Again, most of the photos are archival and non-digitized. Each of the photos presents different aspects of the custom, and those that are stored in different museums can also be used. For example, photos of costumes used in the area of Nestinarstvo.

Another important aspect is audio recordings - of the prayers in front of the icons, of the fire-dancing melodies, the rhythms of the drum, the shout of dancing on embers, and so on. Again we have a large amount of records to use and again digital and analog.

The other important aspect is the video recording of the ritual itself. Naturally, it lasts for hours, and in some places where it takes place it lasts up to 2 days. Then if a continuous recording is made, it will be very difficult to trace the whole custom. If a user wants to view a specific aspect of the custom, it will be very difficult. Moreover, most of the available videos of the custom are only on individual elements. Examination of these elements alone does not give a complete picture of the event.

We see how comprehensive the presentation of this custom should be, which we are all accustomed to say we know or have heard of. But that's not all. So far we have a set of texts, images, audio and video. All these elements should be connected and their roles and location in the custom assigned. References to external sources or related other intangible cultural heritage objects as well as tangible cultural heritage objects may be included.

Namely, even for such a relatively popular custom for the general public, there is a huge amount of unknown matter that should be used.

We will consider our proposed method of digitization starting from the lowest level. Of course, the various intangible cultural objects may not be as comprehensive as the example given, but we should be prepared for them as well.

6.1. Physical location of multimedia objects

The objects themselves that need to be stored will be in some of the previously described formats. According to the experience gained in digitization, annotation and search for objects of different nature, we offer the following formats:

6.1.1. Texts

As we are particularly interested in the information contained in the text, we believe that the use of plain text is appropriate in this case. The advantage is the small file size, format permissions, easy indexing and searching. Using a converter to PDF will also solve the problem of downloading and printing the information.

6.1.2. Images

The most popular format on the Internet, namely JPEG is not very good. It can be used for previews, illustrations, etc., but if we have to stick to the original without losing quality and information, then this format is inappropriate. The disadvantage is that with each processing and compression we lose information every time. Of course, if we have to use a photo or image for informational purposes, then this format is quite sufficient. However, this is not the case if we have to store, for example, the images of the icons used in the given example of Nesrinarstvo. Although they are only an attribute of this custom, they are also cultural heritage objects and should be stored without loss of information so that future researchers can use them.

The PNG format, created as a modern replacement for GIF and widely used on the Internet, has all the advantages we need. The only drawback is that it does not support the CMYK color model, as it is designed for on-screen visualization only. Therefore, if we want professional quality and future use of information, including processing, printing, etc. this format is not suitable.

The TIFF format gives us all the benefits we are looking for. Designed exactly as a format for professional use, it contains all the benefits we need. The shortcomings that we will mention can be eliminated by defining a convention when creating a TIFF file with an image in it.

The first disadvantage is the presence of multiple images in it, which can become a big problem. Readers of TIFF files usually preview the content and mark the various images in it. However, not all software does this. It is possible to search only the first image in the file and when it is found to stop searching for other available Image File Directory (IFD) and bitmap images.

Another important drawback is the presence of tags that may be unrecognizable to the reader of the software used. As already mentioned, the TIFF file can also contain tags that describe it. The first 70 are preserved. However, each group of users can save and use so-called private tags to annotate their content. This is very convenient, but most TIFF image processing software is designed to skip these tags. If we decide to put important information in the file itself and expect it to be available with other software to other users, this will probably not happen. For this reason, we suggest using only standard tags or private tags such as JFIF and GPS, which are maintained by most readers. This could contribute to better presentation of information and more complete use of the image.

We introduce a convention on the basis of which we define TIFF as a format for images without loss of information, as well as TIFF (with JPEG compression), PNG and especially JPEG format for previews and illustrations.

6.1.3. Audio

With the use of audio recordings in the digitization of intangible cultural objects, things are much simpler. MP3 and Ogg formats compress data with information loss, which is not fatal in this case. The events themselves are designed for people, and the human ear is organized to perceive sound waves from 50 to 20,000 hrz. Therefore, cutting all other frequencies will not affect people's perception of sound. Of course, if we need to store sounds, for example, that are intended for zoologists, for example in natural history museums or the like, we should focus on another format. In our case we are guided by the purely human factor and therefore we offer as a main format the much more modern and with more possibilities Ogg Vorbis format, and as auxiliary MP3, and if for some reason we need quality without loss - Wav. Although Ogg maintains text comments that can be structured in metadata, our recommendation is not to resort to this for standardization purposes. MP3 maintains metadata that can be used.

6.1.4. Video

The established as an indisputable standard MPEG 4 satisfies all our requirements. Despite the presence of many recordings in older formats, we recommend using this format as the old recordings should be converted to MPEG 4.

6.2. Physical location of files

In practice, there are 2 methods for developing multimedia libraries.

The first is based on the use of a standard relational database - a well-known model for working with which there are many products and tools. With it, all metadata is arranged as fields in a table. This simplifies search queries, but is extremely inflexible. If we have to change an attribute, it is related to changing the whole database and the logical connections in it. A mixed approach is used, which to some extent eliminates these shortcomings. In it, groups of attributes described in XML are placed in the database tables, which somewhat solves the problem.

The other approach is the so-called native method. In it we have a multimedia server in which we store both the multimedia data and the metadata that describes them. The basic concept for the implementation of this approach is presented below.

We will not dwell in detail on the topic of technical implementation of such a repository of multimedia objects, as the means of implementation are well known.

Figure 1. Library for multimedia objects - basic concept

The proposal is to use appropriate repositories for each type of media - text, images, audio and video. Thus, different tools can be used to create different repositories. In video storage, for example, it is appropriate to use tools that allow streaming.

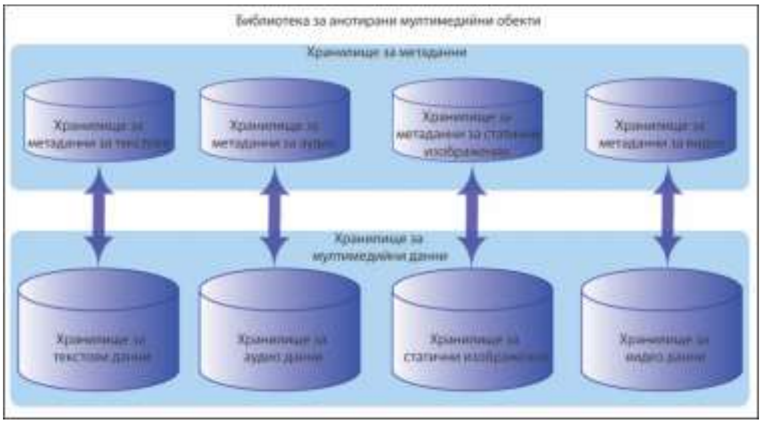

Figure 2. Library for multimedia objects - extended model

7 Text repositories and metadata. Search and storage

Text materials can be both digital and paper. When digitizing materials, our system should have a module for retrieving information from paper. Text documents in digital libraries are often created by digitizing paper documents, importing text files, or from speech. Digitization is the process of converting paper documents electronic. The process usually takes place in several stages. The first is scanning, the second is optical character recognition (OCR) and the last is error checking. We are not talking about indexing and creating metadata yet.

Currently, many libraries have specialized book scanners that perform these operations automatically. The result of the scan is a digital image of each page. If the image includes non-text material, such as diagrams, illustrations, etc. then they should be separated and possibly transferred to the image storage system. When transforming a scanned image into a digital representation, we need to extract the information and present it as text, not a sequence of pixels. Although OCR is automatic, subsequent "manual" inspection and cleaning of the material is always required.

With good error checking software, "manual" work will later be minimized. In our case, due to the main texts that are written in Cyrillic, noise is generated in the system from the very beginning. If we add the proofreaders in Bulgarian, which are not as perfect as those in English, then we have to add a few percents more work that will be done manually.

The problems that are likely to arise are related to proper names, foreign names and words, and special terms. In the case of texts from the archives of libraries or historical museums, more serious problems are possible, arising from the quality of the paper, the fonts used, the quality of the print, etc. Watermarks, notes and any other non-linear texts further degrade the quality of the final result.

The approach to overcoming these problems is as follows. Each text is digitized 2 times by different operators. The results are compared automatically and any differences are removed manually. In this way we can be sure that the end result is as close as possible to the real one. Of course, mistakes are not excluded and it is impossible to avoid them at all. A preliminary analysis and processing of the pages when scanning or a good calibration of the book scanner for the specific material would eliminate many of the problems.

The sound clips from which the text is to be extracted are transcribed using a speech recognition system. The system must be based on pre-prepared acoustics and language to give word accuracy above 50%. It also depends on the style of the speaker. The main method is to look for a semantic connection and check the words in a dictionary.

7.1. Audio libraries

The use of audio files is one of the important contents in the multimedia libraries for intangible cultural heritage. In the world of intangible culture, songs, stories, music occupy a central place. Therefore, special space should be set aside for this type of libraries. A huge audio fund has been accumulated from recordings made through cassette players, tape recorders and other devices. Of interest are also the recordings of stories by performers, eyewitnesses, etc. A possible solution is to transfer these recordings into text in order to use the information in them, but this loses the authenticity of the audio recording. Therefore, we suggest storing the original records as well, using the segmentation and search techniques described here. We believe that the only standard suitable for annotating audio content for our area is the MPEG-7 audio standard.

7.2. Image storage and search

Grading the importance of the data, we come to the part of the library that stores images. One of the good methods for image annotation is the Europena data model (EDM). This model describes the images relatively well, containing elements from Dublin Core, Resource Description Framework (RDF) and others. The model itself is convenient as a standard for exchanging information between libraries within the EU. Although the initiative has fading features and failed to gain the expected popularity, we believe that a module to convert our chosen standard to EDM for compatibility and portal reference purpose, would be an advantage.

The presented analysis shows that the most appropriate standard for annotation needs in multimedia libraries is MPEG-7, as it supports low-level descriptors that can be filled in automatically and through them look for similarities with other images. The main advantage of MPEG-7, which is extremely important in the design, construction, manipulation and operation of multimedia libraries and their expansion, is the support of high-level descriptive descriptors. In the next part of the exposition these descriptors are listed and their main characteristics are presented.

7.3. Video data storage and search

Our main goal is to create and implement a video extraction framework that is modular, scalable and based on MPEG-7. The modularity and use of MPEG-7 will allow the system to be divided into components and information to be exchanged with other systems. The general scheme of a system for searching and retrieving multimedia objects is presented in the following figure.

Figure 3. General scheme of a system for searching and retrieving multimedia objects from MPEG-7 based storage

To ensure well-functioning retrieval methods, the availability of high quality content annotations is essential. This is achieved through content analysis modules. All information about different multimedia data is stored as MPEG-7 annotations. Existing metadata can be transferred to MPEG-7 automatically. The multimedia data together with the corresponding XML for MPEG-7 annotations are stored in a multimedia data store. There are three different types of metadata:

- General metadata as a text annotation
- Low-level metadata such as the color histogram of the image
- Semantic descriptions such as agents, events and relationships

The descriptions used to access content can be classified as follows:

- Storage description: file and encoding formats, image size, image quality; Description and authorship: date and place of creation, title, genre, etc.
- Semantic description: summary of content, events, objects, etc.
- Structural description: key frames - color, texture and movement functions, etc.

Description metadata: author, version, creation date, and last revision

7.4. Multimedia library model

So far, we have looked at standards and proposed solutions on how to create multimedia libraries that store objects annotated in an appropriate way. We also suggested how to perform a semantic search. All this is convenient when creating a simple library, for example for documents, pictures or an audio library. As mentioned in our example, intangible cultural heritage is described by a set of many and varied multimedia objects. So far, we've looked at digital libraries for text, audio, images, and video. Our task is to unite according to some scheme objects from these libraries and thus to present or describe a new complex object which is a unit of the intangible cultural heritage, in the given example Nestinarstvo. The MPEG-21 standard has been identified as the most appropriate method for this purpose.

Let's look at the scheme of a digital object, defined in terms of MPEG-21. As we have already explained, it can describe multimedia objects and other digital objects embedded in it. This makes the structure recursive, extremely flexible and suitable for our case. Each component may be a multimedia one that is described with EDM, MPEG-7 or DC for example. The following scheme is of a digital object represented according to MPEG-21.

Figure 4. Scheme of a digital object according to the MPEG-21 standard

Within our multimedia library, we will introduce the following hierarchy to simplify the definition of a complex object of intangible cultural heritage.

1) First level digital object

It contains only one component of text, image, audio or video type.

The digital object is annotated with MPEG-21, and the content is described using EDM, MPEG-7 or DC for texts.

It does not contain any more components or other digital objects.

Note that the descriptors of the digital object are not specified. This means that they describe the semantics of the resource, while the descriptors in the component that are grouped with the resource describe its technical specification.

Figure 5.Scheme of a first level digital object

In our multimedia library, content sources include objects of the type: text, audio, images, and videos. Users can specify their content requirements through the user environment for request and use. This way we can customize the delivery of content. Of course, in order not to complicate the system, we believe that several types of templates can be used to adhere to.

Content adaptability refers to the many variations through which we can transform content, for example, display the text file as a pdf.