



D4.2.1.1

Standards and Vocabularies for Multimedia Content Description

Author(s): Stina Westman, Pirkko Oittinen

Confidentiality: Consortium

Date and status: July 15th: Version 1.0

This work was supported by TEKES as part of the next Media programme of TIVIT (Finnish Strategic Centre for Science, Technology and Innovation in the field of ICT)

Participants	Name	Organisation
Researchers	Pirkko Oittinen, Stina Westman	Aalto SCI

next Media
www.nextmedia.fi
www.tivit.fi

Executive Summary

Management of multimedia content assets requires a rich metadata set. The metadata is used to manage workflows, offer access to archives, facilitate content repurposing, enable content exchange and build consumer services. Multiple standards are relevant to the description of technical, structural and semantic properties of multimedia assets.

These standards and guidelines for multimedia content description are reviewed in this deliverable with a special focus on content categorization. Content categories, based on program e.g. topic and genre, are central to managing collections of programs and recommending them to users.

Commonly used and promoted metadata schema and metadata vocabularies from various organizations in the media field are presented in this deliverable. Program genre vocabularies by different broadcasters are also compared.

Table of Contents

Executive Summary.....	2
1 Introduction.....	5
2 Multimedia Metadata Standards.....	5
2.1 EBU Standards	6
2.2 IPTC Standards	7
2.3 IRT/BMF	8
2.4 MPEG-7 and MPEG-21	9
2.5 PBCore	9
2.6 PLUS	10
2.7 PRISM/DIM.....	10
2.8 SMEF	11
2.9 SMPTE Metadata Dictionary.....	11
2.10 TV-Anytime	12
3 Multimedia Description Vocabularies.....	12
3.1 EBU Classifications.....	12
3.2 IPTC Newscodes	14
3.3 IPTC Visual Content Controlled Vocabulary	14
3.4 MPEG-7 and TV-Anytime Vocabularies.....	15
3.5 LSCOM.....	15
3.6 Broadcaster Program Vocabularies	16
3.7 Professional and Consumer Vocabularies	17
4 Conclusions.....	18

List of Acronyms and Abbreviations

AMWA	Advanced Media Workflow Association
ARD	Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland
BBC	British Broadcasting Corporation
BMF	Broadcast Metadata Exchange Format
DDL	Description Definition Language
DIM	Digital Image Management
DMS	Descriptive Metadata Scheme
DNG	Digital Negative
DS	Description Scheme
EBU	European Broadcasting Union
EXIF	Exchangeable Image File Format
IEC	International Electrotechnical Commission
IIM	Information Interchange Format
IPTC	International Press and Telecommunications Council
IRT	Institut für Rundfunktechnik
ISO	International Organization for Standardization
JPEG	Joint Photographic Experts Group
LSCOM	Large-Scale Concept Ontology for Multimedia
MPEG	Moving Picture Experts Group
MWG	Metadata Working Group
MXF	Material eXchange Format
PBS	Public Broadcasting Service
PLUS	Picture Licensing Universal System
PRISM	Publishing Requirements for Industry Standard Metadata
RTVE	Corporación Radiotelevisión Española
SAA	Stock Artist Alliance
SMEF	Standard Media Exchange Framework
SMPTE	Society of Motion Picture and Television Engineers
TIFF	Tagged Image File Format
UML	Unified Modeling Language
UPDIG	Universal Photographic Digital Imaging Guidelines
URI	Uniform Resource Identifier
XML	Extensible Markup Language
XMP	Extensible Metadata Platform
YLE	Finnish Broadcasting Company
ZDF	Zweites Deutsches Fernsehen

1 Introduction

The management of multimedia assets requires content description and categorization. Metadata, structured data about the multimedia content, is used for describing content. The descriptions may pertain to the semantic meaning of the content, its structural and technical properties, rights and workflow issues as well as usage information (Pereira et al., 2008).

Metadata typically consists of pre-defined metadata elements with one or more possible values defined by a schema and encoded using an established description language. These metadata descriptions may be global, e.g. tags or categories which provide information about the content as a whole. Descriptions may also be local annotations of regions which provide information about the depicted relations and arrangements of objects. Within a video stream annotations may also have a temporal dimension.

Metadata is used to manage content workflows, preserve access to archived material, facilitate reuse and repurposing of content, enable content exchange between broadcasters and build consumer services. Enriching content with semantic metadata (e.g. keywords, categories) which aims to describe the meaning of the content creates competitive advantage in the changing market where content may be published, distributed and aggregated by anyone.

This deliverable describes standards and vocabularies relevant to multimedia description with a particular focus on content categorization in television production environments. Both standards for still images and video are reviewed. Image file formats (e.g. EXIF¹, Exchangeable Image File Format) or container formats (e.g. MXF², Material eXchange Format) are not reviewed although they can support inclusion of descriptive information. Generic metadata standards (e.g. Dublin Core³) are also not reviewed here, although they are used at the highest abstraction level also for images.

2 Multimedia Metadata Standards

Several standardization bodies have published standards for multimedia metadata. These metadata schemas vary in scope, level of detail and syntax. Figure 1 illustrates the differing scopes of the standards reviewed in this deliverable. Some standards are domain agnostic while others have a strong connection to the television and broadcasting business. Some standards cover the production lifecycle of the content while others focus on the distribution and delivery stage.

¹ http://www.cipa.jp/exifprint/index_e.html

² http://tech.ebu.ch/docs/techreview/trev_291-devlin.pdf

³ <http://dublincore.org/>

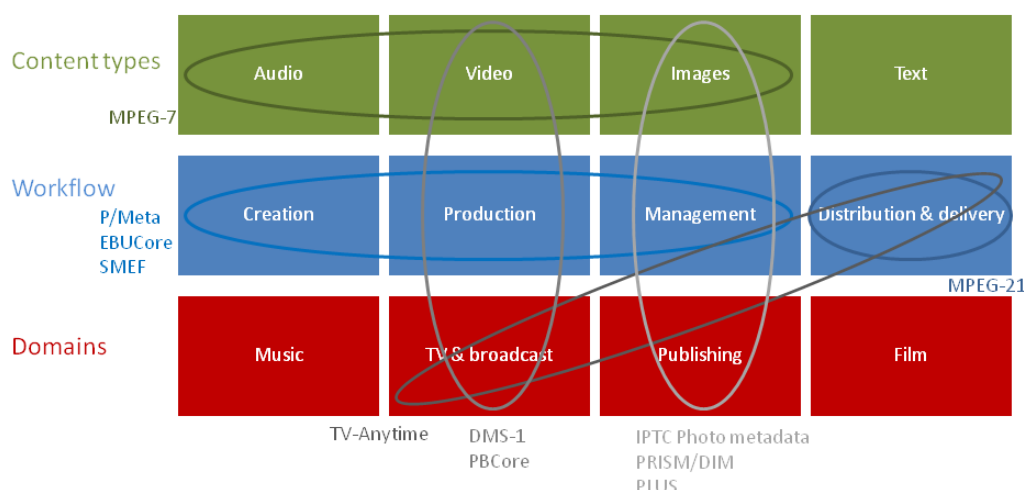


Figure 1. Multimedia metadata standard coverage (adapted from Smith & Schirling, 2006)

2.1 EBU Standards

The European Broadcasting Union (EBU⁴) has published various standards for the management of television and radio program material. It also collaborates on metadata research projects and standardization efforts with other organizations in the field.

P/Meta⁵ was developed by EBU as a semantic metadata schema for descriptive information relating to programs in the broadcasting industry. P/Meta applies to business-to-business content exchange scenarios between content creators, distributors, and archives. Audiovisual material is referred as programs, related program groups, editorially defined program items and continuous media objects of a single media type in the P/Meta schema. The attributes in the schema are considered to be representative of common production practices. Its representation format is XML.

The P/Meta metadata model is hierarchical and all sets and attributes are defined within the standard. The use of classification schemes within P/Meta is defined so that classification schemes must be available as resources on the open Internet via maintained URLs⁶. P/Meta recommends the use of EBU schemes, discussed in section 3.1.

EBU Core Metadata Set (EBUCore⁷) was designed to work as a minimum list of attributes to describe audio and video resources for broadcasting applications including for archives, exchange and publication. EBUCore is based on the Dublin Core standard.

⁴ <http://www.ebu.ch/>

⁵ http://tech.ebu.ch/metadata/p_meta

⁶ http://tech.ebu.ch/docs/tech/tech3336v1_0.pdf

⁷ <http://tech.ebu.ch/lang/en/MetadataEbuCore>

The EBUCore metadata set elements are:

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights
- Version
- Publication History
- Part
- Metadata Provider
- Entity (Contact Details, Organisation Details, Role)
- Type, Status, Format and Date attributes

2.2 IPTC Standards

IPTC (International Press and Telecommunications Council⁸) is a consortium of the world's major news agencies, news publishers and news industry vendors. It develops and maintains technical standards for news exchange that are used by virtually every major news organization in the world. IPTC has published several standards aiming to describe media asset content and provide administrative information. The standards relevant to multimedia content description are the IPTC IIM (Information Interchange Model⁹), IPTC Core Schema and IPTC Extension Schema. The latter two have been defined as XMP (Extensible Metadata Platform) schema, making up the IPTC Photo Metadata Standard¹⁰.

IPTC IIM is a legacy format for text files that describe accompanying media. For example TIFF and JPEG image files include IPTC headers from IIM. Many third-party developers have created software applications that read and write to IPTC headers and the storage locations and structure of metadata varies. The original IIM schema is widely recognized by software products that access metadata, and still useful in cases of tools which cannot read or write the more recently defined XMP schema.

⁸ <http://www.iptc.org/>

⁹ www.iptc.org/IIM/

¹⁰ www.iptc.org/photometadata

IPTC Core is an updated standard which specifies the use of IPTC data within the XMP schema. The standard enables IPTC data to be incorporated via XMP into a wider range of image formats, such as JPEG, TIFF, JPEG2000, DNG and more. XMP supports Unicode text. Unlike IIM, IPTC Core does not have character limits for fields.

The IPTC Photo Metadata standard is defined as the combination of the IPTC Core and the IPTC Extension. These metadata schemas were developed for professional use with a focus on news and stock photos. The Extension fields to provide additional information about the content of the image, (e.g. name or location shown in the image, an organisation or event featured), metadata for cultural heritage photos, (e.g. a title, a creator, a creation date and information about the source of an artwork) and fields to precisely define the licensing terms and the copyrights of a photograph.

These standards are used as de facto standards in the image industry. For example, many fields in common image editing software¹¹ come from IPTC standards. Image agencies have adopted these fields for B2B image exchange. The SAA (Stock Artist Alliance¹²) Photo Metadata Project META has created resources for promoting best practices in utilizing the standards. Among other resources, they have published the Resources Guide to Photo Metadata Fields¹³ which lists and defines the most commonly used metadata fields in key metadata schema.

2.3 IRT/BMF

The Broadcast Metadata Exchange Format (BMF¹⁴) developed by IRT (Institut für Rundfunktechnik¹⁵) consists of a class model describing the meaning of and relationships between information which is relevant and exchanged in television production processes. Based on the BMF UML class model, IRT developed the BMF XML-schema which builds the basis for an XML-based metadata exchange. Controlled vocabularies used by the broadcasters ARD and ZDF are included in the format.

The package MainIndex contains a class for the general description of the programme item, as well as classes for the description of its Theme, depicted Locations, and Persons. Segment-based annotation of thematic and visual content is contained in package Sequence. The package Classification contains classes which make possible the categorisation of the Content of a programme item according to various viewpoints. Categorisation according to genre, intended purpose, presentation form, target group and suitability is supported. The ARD/ZDF vocabulary includes genres such as daily news, economy and sport. Fictional programme content is divided into e.g. television drama, feature films and music clips. Nonfictional content includes for example discussions,

¹¹ http://www.controlledvocabulary.com/imagetdatabases/iptc_core_mapped.pdf

¹² <http://www.stockartistsalliance.org/>

¹³ <http://www.photometadata.org/META-Resources-Field-Guide-to-Metadata>

¹⁴ <http://www.irt.de/bmf/>

¹⁵ <http://www.irt.de/>

interviews, and documentary content. Performance types include e.g. concerts, readings and comedy.

2.4 MPEG-7 and MPEG-21

MPEG-7 (Multimedia Content Description Interface¹⁶), developed by MPEG (Moving Picture Experts Group), is a standard for describing the multimedia content data that supports interpretation of the meaning of the content which can be passed onto, or accessed by, a device. MPEG-7 is a standard framework for describing multiple aspects of multimedia content in several hierarchical levels, including structural and storage information, low-level descriptions of individual objects, high-level abstract descriptions of scenes, and information related to content usage.

The standards consist of multiple parts, including descriptors, a description definition language and system tools. The individual descriptors define the syntax and semantics of metadata elements. Description Schemes (DS) specify the structure and semantics of the relationships between their components, combining descriptors. The Description Definition Language (DDL) defines the syntax of the MPEG-7 Description Tools, based on XML Schema. The System tools support binary coded representation of content for efficient storage and transmission for enhanced management and protection of intellectual property.

The standard includes multiple description schemes, including descriptors for audiovisual content (visual, audio and multimedia descriptors). For example, MPEG-7 Visual Description Tools included in the standard consist of basic structures and Descriptors that cover the following basic visual features: Color, Texture, Shape, Motion, Localization, and Face recognition. Descriptions may focus on the structure of the content (Segment description scheme), or its semantic aspects (Semantic Description scheme). Semantic aspects include the objects, events and concepts depicted in the content while segments can be described with both spatial and temporal properties. MPEG-7 also includes tools for describing videos for summaries and describing collections of videos.

MPEG-21¹⁷ (ISO/IEC 21000) is a framework for exchanging digital content items across devices and networks. It specifies metadata for packaging content, rights information and adaptation to different usage environments. In a sense it complements the MPEG-7 metadata schema which paid almost no attention to publications and rights metadata. Like MPEG-7, MPEG-21 is an XML-based standard.

2.5 PBCore

PBCore¹⁸ is a metadata standard for audiovisual media. It provides a rich set of descriptors for radio and television programs, in both analog and digital formats.

¹⁶ <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>

¹⁷ <http://mpeg.chiariglione.org/standards/mpeg-21/mpeg-21.htm>

¹⁸ <http://pbcore.org/>

Development of PBCore was funded by the Corporation for Public Broadcasting to serve the U.S. public broadcasting community.

The PBCore standard is based on Dublin Core with a number of added elements useful for media. The elements fall into three categories: Content, intellectual property, and instantiation. The content elements describe the intellectual content of the program, while property elements describe its creators and business information. The instantiation elements describe the physical or digital content item. PBCore vocabularies are maintained in the Open Metadata Registry¹⁹. For example, PBCore Genres include entities such as entertainment, dance, French and news.

2.6 PLUS

The Picture Licensing Universal System (PLUS²⁰) is an integrated set of standards for communicating rights information associated with commissioned and stock images. The PLUS standards are developed, approved and maintained by the PLUS Coalition, an international, non-profit association representing publishers, photographers, stock image distributors, museums, libraries and standards bodies, such as UPDIG, IPTC, and IDEAlliance. PLUS licenses can reside in IPTC or XMP metadata, with potential for direct embedding in an image file.

The PLUS Glossary standardizes over 1000 licensing terms, definitions and uses. The PLUS License Data Format is a schema defining a standardized structure for use in communicating an image license summary. It specifies and defines a series of fields used to summarize a license and provides the structure for the PLUS Universal License Statement. The PLUS Media Matrix is a controlled vocabulary for Media Categories, Media Types and **Media Options** as a standardized image licensing menu structure. The Media Matrix may be used to define e.g. that an image is to be published in motion pictures or television (Category), in a documentary program (Type), offered on the internet as a downloadable file, further specified by e.g. image size, placement and duration.

2.7 PRISM/DIM

The Publishing Requirements for Industry Standard Metadata (PRISM²¹) specification for Digital Image Management (DIM) defines a set of XML metadata vocabularies for syndicating, aggregating, post-processing and multi-purposing media content with a focus on magazine content. PRISM uses the Dublin Core and its relation types as the foundation for its metadata. PRISM also recommends practices for using the Dublin Core vocabulary. PRISM extends its metadata element set beyond those selected from Dublin Core in order to specifically allow for fuller description of magazine and journalist content. The PRISM Specification defines a small set of vocabularies for use in characterizing

¹⁹ http://metadataregistry.org/PB_Core

²⁰ www.useplus.org

²¹ <http://www.idealliance.org/specifications/prism/resources/46>

resources. The Namespace includes vocabularies for resource relations, file formats genres, delivery platforms, and rights.

The PRISM Genre element²² describes the genre, or the intellectual content of the resource, separating it from the style of content presentation. Genre is assigned to aid with searches. Sample genres include biographies, cover stories, features, recipes, and fashion shoots. The PRISM Presentation Type²³ which defines the way that a resource presents information or the stereotypical form of the resource. Sample presentation types include chart, infographic, illustration, photomontage and slideshow.

2.8 SMEF

The Standard Media Exchange Framework (SMEF²⁴) was developed by The British Broadcasting Corporation (BBC). Originally it was intended as a data model for the BBC but is now an exchange data model. It consists of a set of definitions for the information required in production, distribution and management of media assets, currently expressed as a data dictionary and set of entity relationship diagrams. Within the SMEF model, a program may be split up into different media objects to be described. The aim is to cover the full content life cycle and to provide an organization independent data model for broadcasting operations. For categorizing content, the editorial description scheme specifies the description context. As the semantics of the standards are open, any description vocabulary may be used.

2.9 SMPTE Metadata Dictionary

The Society of Motion Picture and Television Engineers (SMPTE²⁵) has standardized formats for technical containers for video, including basic metadata. The SMPTE Metadata Element Dictionary RP210²⁶ is a list of structured metadata elements commonly met in television and radio production. Strictly speaking it is not a standard but a collection of data elements that have been generated in the development of a particular set of SMPTE standards.

The current specification defines a tree structure describing elements in six main classes: identification and location, administration, interpretive, parametric, process, relational and spatio-temporal. The entries in the dictionary include nouns, adjectives, verbs etc., making it possible to build rich semantic descriptions of the content. The dictionary is used in conjunction with the Material Exchange Format (MXF) which combines the essence (video stream) with metadata for exchange. Metadata is integrated through Descriptive Metadata Schemes (DMS).

²² <http://prismstandard.org/vocabularies/2.0/genre.xml>

²³ <http://prismstandard.org/vocabularies/2.0/presentationtype.xml>

²⁴ <http://www.bbc.co.uk/guidelines/smf/>

²⁵ <http://www.smpete.org/home/>

²⁶ <http://www.smpete-ra.org/mdd/>

SMPTE standard ST 0380-2004 describes DMS-1 which defines the use of the SMPTE Metadata Element Dictionary in MXF files. DMS-1 has been defined by AMWA (Advanced Media Workflow Association²⁷) as a set of attributes to be associated with audiovisual material in MXF containers. DMS-1 provides a structured data model for descriptive metadata in three classes: production (applies to entire MXF file), scene (describes actions and events within individual scenes) and clip (capture and creation information about the audiovisual clips in the file body). DMS utilizes the native ability of MXF to tie metadata to an essence timeline, not just a header.

2.10 TV-Anytime

TV-Anytime²⁸ is an open standard for metadata describing television programs. The standard was created to support personal recorders and program guides. The TV-Anytime Metadata Specification contains a TVAnytime Usage History Thesaurus, a TV-Anytime Genre Dictionary, and the TV-Anytime Description Schemes, many of which reference MPEG-7 tools. In the context of TV-Anytime, metadata means descriptive data about content, such as program title and synopsis, as well as information about user preferences and history. Metadata may be associated with segments, enabling time-sensitive annotations.

3 Multimedia Description Vocabularies

Content exchange requires standard metadata schema but resolving the values assigned to these fields is also a concern. Metadata values are the actual numbers and terms contained in the metadata fields. Controlled vocabularies are used to achieve interoperability at the level of descriptive terms.

In broadcasting, controlled vocabularies are most often related to terms for genres, subjects and formats. In situations where controlling the values completely is not feasible (e.g. user-centred tagging approaches), guidelines for using metadata fields are given.

Some vocabularies have been discussed in Section 2 jointly with the standards they fall under. In the followings sections, several key vocabularies will be discussed more in-depth.

3.1 EBU Classifications

The EBU has developed several Technical Specifications on Metadata e.g. to support business to business exchange of content (see Section 2.1). These specifications rely to varying extent on the use of common reference vocabularies

²⁷ <http://www.amwa.tv/>

²⁸ <http://www.tv-anytime.org/>

for interoperability purposes. For supporting this, the EBU has published EBU Reference Data & Classification Schemes²⁹.

The widest ranging EBU Classification vocabulary is ESCORT. The focus of the redesigned ESCORT 2007³⁰ is the definition of programme and service concepts and their genres. The ESCORT classification is organised in different dimensions with EBU recommendations given for program:

- Intention (e.g. inform, enrich, entertain, promote, advertise, other)
- Format (e.g. structure/non-fiction, show/entertainment, representation/play/fiction, artistic performance, interactive)
- Content (see EBU content genre classification below)
- Intended Audience/Target Group (e.g. age group, social/regional/minority groups, occupational groups, gender, geographical)
- Origination/Production Mode (e.g. cinema, television, radio, online)
- Content Alerts (e.g. violence, nudity, strobing)

ESCORT thus enables multidimensional classification of a program (see Table 1).

Table 1. Examples of program classifications according to EBU Escort

Programme	Intention	Format	Content	Intended audience	Origination
Pride and Prejudice	Entertain	Presentation, play, fiction	Period drama	General	Cinema
TV News	Inform	Structured	News	General	Television
Football match	Entertain	Commented event	Football	General	Made on location / Live

The EBU Content genre Classification Scheme³¹ scheme is a superset of different classification schemes including in particular TV-Anytime's Content scheme and MPEG-7's Genre scheme (see Section 3.4). It is utilized in ESCORT 2007 and in the EBUCore metadata set. It is up to four levels deep (e.g. non-fiction - daily news, personal/lifestyle/family - fashion).

The EBUCore metadata set also includes a Subject element which refers to the generalized topic that represents the intellectual content, typically expressed by keywords or key phrases. Free text, controlled vocabularies, authorities, or formal classification schemes may be used for subject terms. A longer description (e.g.

²⁹ http://tech.ebu.ch/docs/tech/tech3336v1_0.pdf

³⁰ <http://tech.ebu.ch/docs/tech/tech3322.pdf>

³¹ http://www.ebu.ch/metadata/cs/ebu_ContentGenreCS.xml

abstract, summary, running order) may be provided under the Description element.

3.2 IPTC Newscodes

Within IPTC standards (see Section 2.2), several fields are associated with IPTC-defined taxonomies, NewsCodes³². Especially interesting are the Subject Code, Scene and Genre taxonomies.

The Subject Codes³³ is a three level system for describing content by a well defined set of terms. It currently includes roughly 1400 terms with example codes including architecture, country music, high jump and Christmas. The terms may be broad (e.g. politics) or narrow (100-m hurdles).

The Genre³⁴ code describes the nature, journalistic or intellectual characteristic of a news object, not specifically its content. Example genres include analysis, archive material, interview and raw sound.

The Scene³⁵ code describes the scene of what is covered by the content. Example scene codes include headshot, night scene, action and off-beat.

3.3 IPTC Visual Content Controlled Vocabulary

In addition to metadata standards and NewsCodes, IPTC is developing a Visual Content Controlled Vocabulary. It is to include a set of core terms for describing images in a formal way. The vocabulary is being developed by a work group and only presentations (Saunders, 2010) have been given about the vocabulary.

The goal of the vocabulary is to supply all relevant, unambiguous, and consistent terms for image description to assist image retrieval and reuse. The vocabulary will likely contain terms related to:

- Broad topics and themes
- Objects (orange, teapot)
- Entities (named people, organizations, events)
- Activities (running, swimming, reading)
- Abstract concepts (absence, love, goodness)
- Events (meeting, holiday)
- Groups (political parties, religious groups)
- Description of objects (age, number of people)
- Description of visual aspects (size, color)

³² <http://www.iptc.org/site/NewsCodes/>

³³ <http://cv.iptc.org/newscodes/subjectcode/>

³⁴ <http://cv.iptc.org/newscodes/genre/>

³⁵ <http://cv.iptc.org/newscodes/scene/>

3.4 MPEG-7 and TV-Anytime Vocabularies

The MPEG-7 standard includes multiple classification schemes for audiovisual content. Among these is a Genre Classification Scheme³⁶, used to classify programs based on their content. Within the User Preference descriptions, there is also an element for Classification Preferences.

The TV-Anytime standard documentation has multiple metadata schemas³⁷, including vocabularies³⁸ for e.g. content type, content format (e.g. artistic performance, a magazine show, a cartoon), origination (e.g. live from a studio, a cinema release movie), and atmosphere.

The Content Classification³⁹ contains terms that describe the nature and and/or subject of the programme. The scheme is four levels deep at its deepest (e.g. Non-fiction- Philosophies of life - Non-religious philosophies – Humanism).

The Atmosphere Classification scheme contains terms that convey the psychological or emotional characteristics of a content item. These include values such as chilling, fast-moving, and practical.

The development work of the MPEG-7 and TV-Anytime genre vocabularies coincided with EBU's efforts and there are considerable similarities between the taxonomies. These approaches to genre in broadcast media are hoped to meet the requirements for programme classification in both the professional and consumer domains.

3.5 LSCOM

The LSCOM (Large-Scale Concept Ontology for Multimedia⁴⁰) workshops aimed at defining a standard formal vocabulary for the annotation and retrieval of video. The workshops brought together researchers in multimedia learning, information retrieval, computational linguistics, library science, and knowledge representation, as well as user communities such as intelligence agencies and broadcasters. Concepts were evaluated based on their utility in video retrieval, the feasibility of their detection and their common occurrence in actual video data sets.

The current concept list⁴¹, expanded through the Cyc knowledge base includes more than 3000 concepts. The concepts include proper nouns (e.g. George Bush), nouns (e.g. man), action labels (e.g. people marching), and affective terms (e.g. fun). The LSCOM concepts are employed in the TRECVID evaluation campaign⁴² for video retrieval. In the campaign, academic teams build automatic detectors for the concepts. In order to reduce the effort in replicating similar

³⁶ <http://www.ebu.ch/metadata/cs/mpeg/mpeg7-genreCS.xml>

³⁷ http://tech.ebu.ch/docs/metadata/ts_1028220301v010601p.pdf

³⁸ <http://www.ebu.ch/metadata/cs/tva/>

³⁹ <http://www.ebu.ch/metadata/cs/tva/ContentCS.xml>

⁴⁰ <http://www.lsc.com.org/>

⁴¹ <http://www.lsc.com.org/concept.htm>

⁴² <http://trecvid.nist.gov/>

baseline systems, several research teams have made their baseline detectors to the research community^{43,44,45}.

The first version of the LSCOM annotations consists of labels for 449 visual concepts. These concepts have been annotated based on human judgments of relevance for keyframes extracted from the videos. Revised Event/Activity annotations⁴⁶ based on video were conducted on 24 concepts, which contained a temporal component (e.g. car crash, people crying).

3.6 Broadcaster Program Vocabularies

For the purpose of comparing categorization vocabularies, genre listings were obtained from various international broadcasters (BBC⁴⁷, PBS⁴⁸, RTVE⁴⁹, YLE⁵⁰, ZDF⁵¹) through their websites. Additionally, the main competition categories of the Banff World Media Festival⁵² were examined. The categories were compared against each other and the Genre vocabulary⁵³ by EBU.

Most published categorizations were on a single level and based on program genres. They included from nine to eleven main program categories. Common genres found in all vocabularies included sports, entertainment, children's programming, news, drama and music. Regional differences were evident in some genres, such as Crime in the ZDF taxonomy, Weather in the BBC ontology, and Movies in the RTVE vocabulary. The BBC categorized programmes by both their genre (e.g. factual) and their format (e.g. documentary). Most taxonomies were utilized to aid end-user searching and browsing of program content online. The PBS taxonomy was however intended as a common metadata layer for various broadcasters for pushing video into pbs.org services, and it had several genres not present in other categorizations, such as health, parents and technology. The EBU and Banff categorizations both included an entry for interactive programming, not present in the broadcaster's taxonomies.

There is little public documentation on how broadcasters utilize these program vocabularies in their workflows or in services offered to viewers. The BBC uses a programme's ontology⁵⁴ to describe their content in a controlled manner. The ontology provides web identifiers for broadcasting content concepts such as brand, series, and episode. The ontology captures categorical information about programmes, and relations between such categories. Based on these ontological descriptions the BBC is able to create links between programs from e.g. different

⁴³ <http://www.science.uva.nl/research/mediamill/challenge/>

⁴⁴ <http://www.ee.columbia.edu/ln/dvmm/columbia374/>

⁴⁵ <http://vireo.cs.cityu.edu.hk/research/vireo374/>

⁴⁶ http://www.ee.columbia.edu/ln/dvmm/lscm/LSCOM_event.pdf

⁴⁷ <http://www.bbc.co.uk/programmes>

⁴⁸ http://www.pbs.org/modules/docs/merlin/PBS_Topic_Taxonomy.pdf

⁴⁹ <http://www.rtve.es/alacarta/>

⁵⁰ <http://areena.yle.fi/>

⁵¹ <http://www.zdf.de/ZDFdeEPG/include-extended-searchform.jsp>

⁵² <http://www.banff2010.com/2011.program.categories.php>

⁵³ http://www.ebu.ch/metadata/cs/ebu_ContentGenreCS.xml

⁵⁴ <http://www.bbc.co.uk/ontologies/programmes/2009-09-07.shtml>

media on their website based on permanent concept URIs (see. e.g. <http://www.bbc.co.uk/nature/life/Tiger>). They use DBpedia as a controlled vocabulary, utilizing Linked Data to combine the classification efforts of different teams and domains.

3.7 Professional and Consumer Vocabularies

Several resources exist for semi-professional and freelance photographers as well as for consumers to guide the description of their images.

The Guide to Photo Metadata Fields⁵⁵, published by the Photo Metadata Project, provides a valuable resource for all photo metadata users. It lists and defines the most commonly used fields in key metadata schema and in popular software. Based on photo metadata standards, the Controlled Vocabulary Keyword Catalog⁵⁶ is a software tool for image keywording, interoperable with common image editing and management software. It contains approximately 11,000 terms organized in a hierarchical structure and includes the full IPTC Subject, Scene and Genre codes. The top level of the keyword hierarchy is shown in Figure 2.

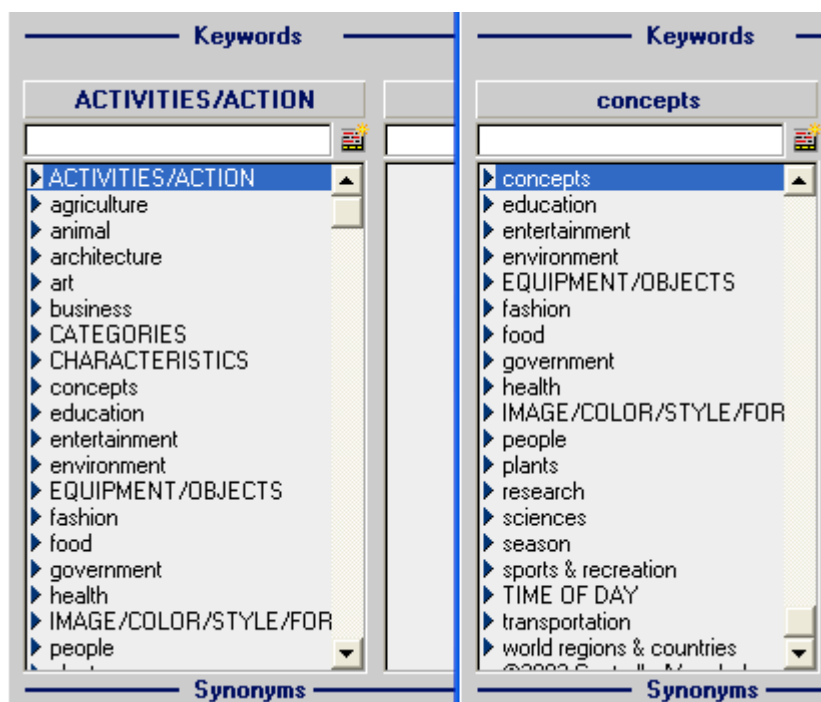


Figure 2. User interface view of the Controlled Vocabulary Keyword Catalog

The Metadata Working Group (MWG) has published Guidelines for Handling Image Metadata⁵⁷ which addresses issues of digital imaging metadata for typical consumers. The guidelines utilize and extend existing standards to address the key organizational metadata questions that most consumers have:

⁵⁵ <http://www.photometadata.org/META-Resources-Field-Guide-to-Metadata>

⁵⁶ <http://www.controlledvocabulary.com/>

⁵⁷ http://www.metadataworkinggroup.org/pdf/mwg_guidance.pdf

- Who is involved with this image? (who took it, who owns it, who's in it?)
- What is interesting about this image?
- Where is this image from?
- When was this image created or modified?

The guidelines include models of usage and data, best practices for including metadata in consumer images based on standards and delineation of new metadata fields. The guidelines recommend the use of hierarchical keywords for consumer images. They also suggest implementing image categorization based on common cases in consumer imagery. Image categories should be created based on content (who, what, where) (e.g. my favourite flower shots, family photos), based on time (when) (e.g. summer 2009 photos), based on resources or processes (e.g. scans from Grandpa) and owner- or user-based (Bobby's photos).

4 Conclusions

Multimedia metadata models have been created for both still images (IPTC Photo metadata), video (e.g. TV-Anytime), and multiple media types (e.g. MPEG-7). Issues with the integration of different media types still exist, as there exists no common metadata model for e.g. photos, videos and associated stories or published articles). The scope of the metadata element set ranged from full coverage of identification, description, technical and rights metadata (e.g. P/Meta) to only addressing a subset of these (e.g. rights metadata in MPEG-21 or PLUS). The consensus in the domain seems to be that multiple standards are inevitable due to the scope of the field. Standardization efforts are nowadays collaborations between several standardization bodies and organizations.

Some metadata models were originally meant for internal use (e.g. BBC SMEF) but all have seemed to develop into data exchange models. There is however little context retained in the presentation of multimedia content metadata which would aid its interpretation amongst different user groups. Tools for multimedia authoring which would include functionalities content description and packaging with essence would be highly beneficial. One of the key drivers of multimedia content description is the need to fulfil queries for multimedia content. MPEG has recently started to develop the MPEG Query Format⁵⁸ which aims to provide a standard format for content requests sent to the server, subsequent responses and additional tools for query management. The ultimate goal is to provide a standardized way to accept and respond to user requests in multimedia search within media applications and services.

Hierarchical models (e.g. MPEG-7) were more common than flat ones (EBUCore). In these models, suitable metadata elements and values for differing levels of granularity in the description need to be specified. The association of metadata to for example different segments of a program, allows restructuring and repurposing of the original media stream to generate new consumption and navigation modes, e.g. video summaries. Also with still images, region-based

⁵⁸ <http://www.mpegqueryformat.org/>

metadata enables targeted searching and requires extra attention throughout image lifecycle and in operations such as resizing, rotation or cropping.

When the goal is to target audience with content, content models need to be matched with user models. In the TV-Anytime standard for example, the atmosphere of the program is classified, as is the mood of the viewer. These types of contextual and affective features - currently accounted for by some of the standards reviewed - add a new dimension to multimedia description and its applications. On the other hand, low-level content descriptors based on basic visual and audio features are still underutilized in systems. Thus the benefits of for example MPEG-7 as an interoperable framework for multimedia description may yet be largely unrealized.

Content categorization was addressed by most standards reviewed. The genre vocabularies have been developed in collaboration, and show many similarities. Also the currently used genre taxonomies from broadcasters had significant overlap. Most categorizations were based on the topics the programs dealt with. However, also categorization based on program format and atmosphere was discovered. These multifaceted categorizations deserve more attention as they would facilitate richer content searches and discovery. Together with structured, hierarchical descriptions of program content, they would enable new modes of adaptive interaction with content.

REFERENCES

Pereira, F., Vetro, A., & Sikora, T. (2008). Multimedia Retrieval and Delivery: Essential Metadata Challenges and Standards. *Proceedings of the IEEE* 96(4): 721-744.

Saunders, S. (2010). IPTC Controlled Vocabulary. Presentation at Photo Metadata Conference 2010, June 9th 2010 Dublin, Ireland. Available at: <http://www.phmdc.org/phmdc2010/PhMdC2010--SSaunders-ImageCV.pdf>

Smith, J. R., & Schirling, P. (2006). Metadata Standards Roundup. *IEEE MultiMedia* 13(2): 84-88.