

Flexible Services

Social Video Metadata Formats

Document Title:	Metadata Formats of Social Video
Date of delivery	
Editor	Chenyuan Peng
Version and	version 1.0
Date Last Change	27.9.2010
File:	

Table of Contents

Table of Contents	1
List of Tables	1
1. Metadata Formats	2
Metadata Requirements	2
Metadata Attributes	3
Metadata Encoding and Delivery	8
<i>Example</i>	9
1.1.1. General Metadata Issues	9
Metadata Sources	9
Automatic Enrichment of Metadata	10
Utilization of Metadata with or without the Video, and Vice Versa	11
References	Error! Bookmark not defined.

List of Tables

1. Table: System Requirements for Social Video	2
2. Table: Service Requirements for Social Video	3
3. Table: The Base Social Video Metadata Elements	4
4. Table: The Social Video Contextual Metadata Elements	5
5. Table: The Social Video Audiovisual Annotation Metadata Elements	6
6. Table: The Social Video User Profile Metadata Elements	6
7. Table: The Social Video Event Metadata Elements	6
8. Table: The Social Video Social Networking Metadata Elements	7
9. Table: The Social Video Enhanced TV Guide Metadata Elements	8

Metadata Formats

Having rich and value-added metadata for user generated mobile video content provides many opportunities for storing, managing, searching, sharing and tracking resources. However, handling metadata can be a very time consuming and resource heavy activity. It requires more descriptive data about the content regardless of the format and the location where it is watched.

Objective is to provide a metadata specifications based on a set of common standards. The appendix specifies metadata attributes, encoding format and delivery mechanism according to the requirements. The initiative was to use the concept of linked open data.

Metadata Requirements

Tables 1 (System Requirements) and 2 (Service Requirements) list the metadata requirements for social video.

System Requirement	Description
Multi-channel environment	CableTV, IPTV, mobileTV, localTV, 3G, Internet
Interoperate with other metadata repositories	Series, movies and documentaries, and file archive metadata management
Collective video production	Innovative use of video metadata and voice tagging extracted from mobile device. Semi-automatically edit and select several sources of information. Fusing multiple media sources into a single one
Enable data sharing, machine-interpretable and machine-creatable	For automatic metadata generation and automated data management, efficiently use metadata in information filtering, retrieval, and storage
Interactive participation services between user and system, and sharing information through social networking	End user's contribution, commenting, and even correction of the metadata associated with content. Sharing user generated news via such as Facebook and twitter
Collect experience based on information and enrich video content	Metadata as an abstraction of sensing, experiencing, like behavior, social judgement, contextual awareness, and personality
Enable semantic search	Social metadata should assist Semantic search, including sharing and retrieving information
Metadata Synchronization	Synchronize with timed content

1. Table: System Requirements for Social Video

Service Requirement	Description
TV program guide	Detailed program descriptions as well as mechanisms to search content in depth, finding not only the top level information, but dialog and scenes contained within the video. More sophisticated use of this metadata allows for better recommendations of content
Dynamic Advertising	Where appropriate ads are displayed based on content of the video
Online video from and for social networks	Users expect to be able to contribute, comment on, correct metadata associated with content and upload video on video sharing sites. Large-scale social network communities such as Napster, Facebook, Twitter, and YouTube where millions of users form a dynamically-changing infrastructure to share content. Therefore, metadata should ensure common definitions for basic information such as title, date, author and language and free tags, which to be used in video upload forms and video feeds of data coming from each site.
Social or collaborative tagging during live event such as sports and concert	Social metadata enables a true interactive and social experience with the end user. Many users add metadata to a resource in the form of keywords and even time-based tags to shared content, so that they can not only categorize information for themselves, they can also browse the information categorized by others
Context aware retrieval of information, personalized event and news filtering, location based services	Mobile usage contextual and sensing metadata are required for context aware retrieval of information using metadata
Speech recognition for tagging the videos live at a location	Using speech to text functionality to create specific user generated content metadata. Speech recognition systems should be able to generate searchable text that is indexed to time code on the recorded video. Results include more efficient and enhanced sorting, storing and searching of video
Tourist and culture info guide	Handling and filtering of metadata can facilitate the creation of cross cultural narratives. Metadata should be able to meet the theoretical and practical challenges of understanding mediated communication in a specific cultural context
Audiovisual annotation application	Subjective comments, notes, explanations or external remarks that can be attached to a video. When a user retrieves a video, they can also download the annotations attached to it

2. Table: Service Requirements for Social Video

Metadata Attributes

We used the concept of metadata "bundle", that is, in addition to the base social video metadata element set, other optional metadata sets can be combined for

different scenarios. However, the base metadata is the minimum requirement. Also each element should be used together with classes when necessary, e.g. SIOC¹ (Semantically-Interlinked Online Communities) and FOAF² (Friend-of-a-Friend). The following tables are social video metadata element sets. Each table is a bundle of metadata element. As mentioned above, the elements can be modified.

Attribute	Vocabulary	Namespace	Description
title	DCMES	dc	The title of the video file
creator	DCMES	dc	The creator of the video file
date	DCMES	dc	The date that the video was uploaded. This value is automatically added when the metadata is initially entered into the service. DCMI Recommended best practice is to use the ISO 8601 profile for date format: YYYY-MM-DD
identifier	DCMES	dc	The URL of the video file
subject	DCMES	dc	A list of keywords describing the content of the video. This is where subject words of a formal classification scheme would be used. User supplied tags/free tags for possible open input area
format	DCMES	dc	The format of the video file. DCMI Recommended best practice is to use MIME type information
rights	DCMES	dc	A URL to a text-based description of the video's licensing terms
description	DCMES	dc	Provides a text-based description of the video
relation	DCMES	dc	The relationship of this resource to another resource. This element should specify what the relationship is (using the type attribute)
source	DCMES	dc	Objects, either electronic or printed, from which this resource was derived
type	DCMES	dc	The nature or genre of the content of the resource. Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary
extent	DCMI	dcterms	The size or duration of the resource

3. Table: The Base Social Video Metadata Elements

The Simple DCMES³ (Dublin Core Metadata Element Set) consists of 15 metadata elements, from which we found main part suitable for us. The additional terms are

¹ <http://rdfs.org/sioc/spec/>

² <http://xmlns.com/foaf/spec/>

³ <http://dublincore.org/documents/dces/>

identified, generally in working groups of the Dublin Core Metadata Initiative (DCMI⁴).

Attribute	Vocabulary	Namespace	Description
position	GeoTag ⁵	geo	Raw latitude;longitude position coordinates for location information about the current location of the device. To extract more high-level meaning, one can use a publicly available Geographic Information System (GIS) resource to determine the address, e.g., street names, city, state, and landmarks, etc
placename	GeoTag	geo	The name of the place
region	GeoTag	geo	The name of the region, country
weather, temperature, humidity	SensorML ⁶	swe	Current environmental conditions including weather, temperature, and air conditions
atDateTime/ created	Timeline Ontology ⁷ / DCMI	dl/dcterms	The current system time of the content consumption

4. Table: The Social Video Contextual Metadata Elements

Attribute	Vocabulary	Namespace	Description
video	SearchMonkey Media ⁸	media	A video object. Use this property to relate a document to a Video object embedded in it or in general to relate an object to an video representation of that object
thumbnail	SearchMonkey Media	media	A thumbnail image, typically of small dimensions
article	SearchMonkey Media	media	An article
text	SearchMonkey Media/Media RSS ⁹	media/mrss	Allows the inclusion of a text transcript, closed captioning, or lyrics of the media content. Many of these elements are permitted to provide a time series of text
fileSize	SearchMonkey Media	media	The size of a resource in bytes
views	SearchMonkey Media	media	The number of times an item has been viewed

⁴ <http://dublincore.org/>

⁵ <http://geotags.com/geo/geotags2.html>

⁶ <http://www.opengeospatial.org/standards/sensorml>

⁷ <http://motools.sourceforge.net/timeline/timeline.html>

⁸ <http://developer.yahoo.com/searchmonkey/smguide/searchmonkey-media.html>

⁹ <http://video.search.yahoo.com/mrss>

download	Media RDF	media	A link to a complete video file representation of the video recording
keywords	Media RSS	mrss	Highly relevant keywords describing the media object with typically a maximum of ten words
category	Media RSS	mrss	Allows a taxonomy to be set that gives an indication of the type of media content
comments	Media RSS	mrss	Allows inclusion of all the comments media object has received
scene	Media RSS	mrss	Specify various scenes within a media object

5. Table: The Social Video Audiovisual Annotation Metadata Elements

Attribute	Vocabulary	Namespace	Description
firstname	FOAF	foaf	The first name of the device user
surname	FOAF	foaf	The family name of the device user
language	DCMES	dc	The preferred language
gender	FOAF	foaf	The gender of the user
accountName	FOAF	foaf	The account name of a video service site
accountServiceHomepage	FOAF	foaf	The URL of a video service site
homepage	FOAF	foaf	The home page of the device user
img	FOAF	foaf	The photo of the device user
email	FOAF	foaf	The email address of the device user
interest	FOAF	foaf	The interests of the device user

6. Table: The Social Video User Profile Metadata Elements

Attribute	Vocabulary	Namespace	Description
dtstart	vevent	cal	start date and time
dtend	vevent	cal	end date and time
attendee	vevent	cal	name of attendee
summary	vevent	cal	event summary
location	vevent	cal	event location
description	vevent	cal	general description of event
organizer	vevent	cal	event organizer
comment	vevent	cal	comment

7. Table: The Social Video Event Metadata Elements

Class	Vocabulary	Namespace	Description
Community	SIOC	sioc	A high-level concept that defines an online community and what it consists of
Container	SIOC	sioc	An area in which content Items are contained
Forum	SIOC	sioc	A discussion area on which Posts or entries are made
Item	SIOC	sioc	Something which can be in a Container
Post	SIOC	sioc	An article or message that can be posted to a Forum
Role	SIOC	sioc	A function of a User within a scope of a particular Forum, Site, etc
Site	SIOC	sioc	Can be the location of an online community or set of communities, with Users and Usergroups creating Items in a set of Containers. It can be thought of as a web-accessible data Space
Space	SIOC	sioc	Is a place where data resides, e.g. on a website, desktop, fileshare, etc.
Thread	SIOC	sioc	A container for a series of threaded discussion Posts or Items
User	SIOC	sioc	A User account in an online community site
Usergroup	SIOC	sioc	A set of User accounts whose owners have a common purpose or interest. Can be used for access control purposes

8. Table: The Social Video Social Networking Metadata Elements

Class/Attribute	Ontology	Namespace	Description
broadcaster	Programmes ontology ¹⁰	po	An organization responsible of some broadcasting services. It can hold a set of services
service	Programmes ontology	po	A broadcasting service. Instances of this concept include BBC Radio Wales, BBC Radio 4, BBC News, etc
frequency	Programmes ontology	po	Associates a channel to its frequency
channel	Programmes ontology	po	A physical channel on which a broadcast occurs. A single outlet or service can be associated with multiple channels
programme	Programmes ontology	po	A TV program
programmitem	Programmes ontology	po	A programme that can have versions, and as such can be broadcast or made available on-demand, e.g. a clip or an episode
IPStream	Programmes ontology	po	An IP stream for DVB-H
tag	Programmes ontology	po	Associates an episode to a particular tag
microsite	Programmes ontology	po	Associates a programme to its microsite. E.g. , http://www.bbc.co.co.uk/programmes/b00fm04s
synopsis	Programmes ontology	po	The synopsis of a series, brand or episode

9. Table: The Social Video Enhanced TV Guide Metadata Elements

Metadata Encoding and Delivery

To meet the social video requirements, we utilize RDFa (Resource Description Framework - in - attributes). RDFa+ XHTML RDFa-in-xhtml. RDFa - a mark-up language for embedding RDF data in Web pages to deploy metadata and reuse existing multimedia metadata vocabularies to enable tools to convert, validate, and exchange the metadata.

¹⁰ <http://www.bbc.co.uk/ontologies/programmes/2009-09-07.shtml>

Example

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
  "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  version="XHTML+RDFa 1.0" xml:lang="en">
<head>
  <title>Peng's Home Page</title>
  <base href="http://example.org/peng/" />
  <meta property="dc:creator" content="chengyuan peng" />
</head>
<body>
  <h1>Peng's Home Page</h1>
  <p>My name is <span property="foaf:nick">Peng</span> and I like
    <a href="http://dbtune.org/musicbrainz/" rel="foaf:interest"
      xml:lang="en">music</a>.
  </p>
  <p>
    My <span rel="foaf:interest" resource="urn:ISBN:0752820907">favorite
    book</span> is the inspiring <span about="urn:ISBN:0752820907"><cite
    property="dc:title">Weaving the Web</cite> by
    <span property="dc:creator">Tim Berners-Lee</span></span>
  </p>
</body>
</html>

```

1.1.1. General Metadata Issues

Metadata is needed to describe the video so that it can be searched and recommended, and therefore found. In the project, we researched and developed automatic metadata detection and speech recognition to get metadata. The process is described in more detail in pilot "Suitcase full of Stories" in the Final Report of Social Video.

Metadata Sources

Mobile phones are a special case with their GPS data possibility. The phone makes it possible to get metadata by automatic detection, textual information and spoken information.

User-generated metadata was mentioned in chapter 2.1.2. See more about getting data from users for recommendations and giving recommendations to users in the chapter (Recommendations) in the Final Report of Social Video..

Automatic Enrichment of Metadata

The automatic metadata can be got from the data saved with in the video by a recording device or conclude from existing metadata or video's audio visual content.

1) *Metadata from Video File*

Mobile phones have e.g. GPS data possibility that can be utilized as a basic for location metadata.

2) *Metadata from Video's Textual Description*

Some videos are without e.g. the video capture time and GPS location metadata. Because those metadata fields are important e.g. in location and event base services, we studied the problem of enriching existing metadata in the event detection case (see chapter 3.8 in the Final Report of Social Video.).

In the event detection case we mined the textual descriptions given in the metadata to find these pieces of information. Regular expressions were used for finding date values within the text, and based on them, record time information was figured out. For the GPS location, we separated the text into words and gave them to a geo-coding service. Then, location values are obtained from the system. Metadata that was gathered this way is, based on our studies, not completely reliable, but it is very useful when used in cooperation with our clustering system.

Also, we utilized techniques that allow us to extract representative keywords (text) for sets of videos. In our case, this means that we select keywords from the videos of one particular event in such a way that the resulting words represent the event well.

3) *Metadata Recognition by Speech Recognition*

The project gave evidence to the assumption, that using speech recognition software to extract spoken content from the audio track of video material is an exciting and rewarding method for acquiring rich metadata for search, tagging, classification, etc. as long as the video material contains speech of decent quality and without too loud background noise. And even the fact that the speech recognition software does not detect any interpretable speech is a fact of value.

Lingsoft refined and customized its Finnish STT (speech-to-text) application suite during and within the project to be used for:

- extracting detected spoken fragments to text
- normalize and filter the recognized text to tags
- attaching time stamps to (the beginning of) recognized word

The project showed that it is possible to use speech recognition for various information retrieval purposes, if only the spoken content can be isolated well-enough. A case for further elaboration is to look at what Lingsoft's core linguistic technologies could provide for post-processing the recognized textual content.

Lingsoft's morphology-aware spelling and grammar checkers, and their fuzzy-logic suggestion mechanisms, can be used for validation and filtering of adequate words, expressions and sentences.

4) *Metadata from Video Content*

Video content analyses are possible, but not yet a valid commercial application, mainly because of the requirements they place upon processing power and data transfer infrastructure.

Utilization of Metadata with or without the Video, and Vice Versa

Because of the current copyright restrictions, some video content cannot be directly shared between all users. This is a limitation, not an absolute restriction! There is demand for a way to share metadata with the users who have access to their own recording copy of the same program (see Adding Social Elements to Time-Shift TV in chapter 3.7 in the Final Report of Social Video). The easiest way to meet this demand is to link additional, later generated metadata to common metadata that identifies the relevant content. This leads to a need to filter the metadata content in the user interface by access, in order to prevent crowding the common public with information irrelevant to them.

Additionally, the same requirement opens up a possibility to view some of the metadata separately from the video content (example: public recommendations and reviews).

Still, there needs to be an option to view the content without the metadata, because it is not always needed, and internet connections are often limited.