



## Cosa sono i Metadata: la parola all'esperto

contributo di Frediano Di Carlo, Consulente per la sicurezza e le tecnologie



### Introduzione

Il mercato della videosorveglianza continua inarrestabile la propria crescita e le telecamere aumentano nelle nostre città di giorno in giorno e, con esse, l'ingente mole di dati dai quali estrarre informazioni d'interesse pubblico e privato.

Di recente, abbiamo assistito, tramite i mezzi d'informazione, a incoraggianti risultati investigativi delle FF.OO., anche per mezzo della videosorveglianza.

Si pensi al caso di **Yara Gambirasio**, nel quale una prova importante è costituita dalle riprese di più passaggi del furgone dell'inquisito nelle zone d'interesse delle indagini; o al caso del piccolo **Loris Stival**, nel quale una prova determinante a carico della madre, presunta assassina, è data dalla ricostruzione degli spostamenti effettuati la mattina

della scomparsa del piccolo, grazie all'analisi di diversi filmati di telecamere della zona.



Eppure, ci sono voluti giorni per raggiungere tali risultati, tempi di certo inadeguati in caso d'indagini dove il fattore tempo è determinante. D'altra parte, a pensarci bene, non potrebbe essere altrimenti: ancora oggi, per la ricerca d'informazioni dalla videosorveglianza si utilizza la stessa fonte degli esordi di tali sistemi, la registrazione video. Ultimamente, è corredata da alcuni strumenti per velocizzare le ricerche, ma sempre legati a eventi predeterminati d'interesse e non, come negli esempi precedenti, a eventi che fuori dello specifico contesto rappresentano la semplice normalità.

Questa situazione è destinata purtroppo a peggiorare, al crescere del numero di telecamere e della complessità degli eventi da ricercare.

Per venire incontro a tali inconvenienti e arricchire il contenuto informativo delle scene riprese, alcuni mesi fa ON-VIF ha rilasciato delle specifiche sulla gestione del Metadata. Il presente articolo ne descrive i contenuti, l'utilizzo e, soprattutto, le potenzialità destinate a mutare profondamente il concetto di Ricerca degli Eventi.

### Etimologia

Il termine **Metadata** deriva in parte dal Greco **meta** (μετα), che significa *con, oltre, dopo*, e dal Latino **data** (plurale di datum), che significa *dati, informazioni*. Il significato è dunque "con le informazioni", "oltre le informazioni"; una definizione adatta allo scopo è **Ulteriori Informazioni (sui Dati)**.

Dato: Libro	Metadata: Scheda Bibliografica
	

Un esempio per chiarire: se il nostro dato è rappresentato da un Libro, possibili altre informazioni sullo stesso, ossia un set di metadata, sono rappresentate dalla Scheda Bibliografica.

Sovente si utilizza anche il termine TAG, ma si tratta di un'espressione impropria, essendo esso uno degli elementi costituenti i metadata, per completare l'esempio precedente:



### Attività di Standardizzazione

Non è del tutto corretto parlare di **standard** per i metadata, essendo tali informazioni le più disparate possibile e ciascuna con propria specificità.

Se, comunque, si esegue una ricerca in Rete utilizzando i termini *metadata* e *standard*, emergono preponderanti risultati legati allo standard ISO 19115, definito "Geographic Information – Metadata"; non meraviglia scoprire che la maggior parte delle attività degli enti normatori sono rivolte alla standardizzazione delle informazioni di geolocalizzazione, visto il ruolo primario ricoperto dalle stesse nell'attuale vita quotidiana.

Per trovare informazioni specifiche su attività di standardizzazione d'informazioni legate ai filmati ovvero, più generalmente ai contenuti multimediali, è necessario risalire agli anni a cavallo tra la fine del secolo scorso e l'inizio dell'attuale, quando fu costituito il gruppo **MPEG** (Motion Picture Expert Group), il cui leader riconosciuto è il nostro **Leonardo Chiariglione** (<http://www.chiariglione.org/>), tuttora parecchio attivo nello specifico settore. Il gruppo mise a punto i seguenti standard, pesantemente entrati nella nostra vita quotidiana:

**MPEG-1:** ISO/IEC 11172 (1993) – «Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s»

- come supporto per la nuova codifica video fu introdotto il "**Video CD**";
- la nuova codifica audio prese il nome di **mp3**, contrazione di "MPEG1 Layer 3", derivante dal fatto che la stessa era specificata dal terzo documento (layer) dello standard.

**MPEG-2:** ISO/IEC 13818 (1994) – «Generic coding of moving pictures and associated audio information»

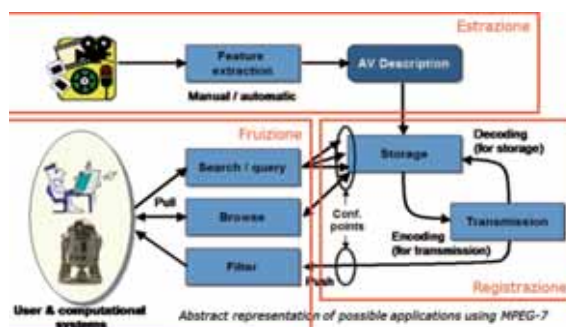
- come supporto per la nuova codifica video fu introdotto il "**DVD**", *Digital Versatile Disk* o *Digital Video Disk*;
- fu introdotta una nuova codifica audio che prese il nome di **aac**, *Advanced Audio Codec*.

**MPEG-4<sup>(1)</sup>:** ISO/IEC 14496 (1999) – «Coding of audio-visual objects»

- come supporto per la nuova codifica video fu introdotto il "**Blue Ray Disk**";
- la codifica audio rimase la precedente **aac**;
- nel 2003 fu introdotta, con il Layer 10, una nuova codifica video nota con il termine **H.264** (frutto del lavoro del Gruppo Misto costituito da MPEG e ITU-T, *International Telecommunication Union*);
- sempre nel 2003 il Layer 14 introdusse il formato file **mp4**.

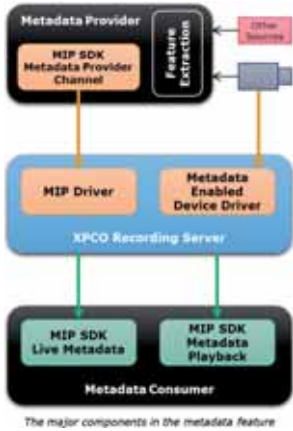
Nel 2002 fu promulgato lo standard **MPEG-7** (ISO/IEC 15938) «Multimedia Content Description Interface», che rappresenta lo standard di Metadatazione dei contenuti Multimediali applicabile ai precedenti MPEG-4, 2, 1<sup>(2)</sup>.

La figura a lato, tratta dalla documentazione dello standard, illustra schematicamente il meccanismo di estrazione, salvataggio e successivo utilizzo, dei metadata.



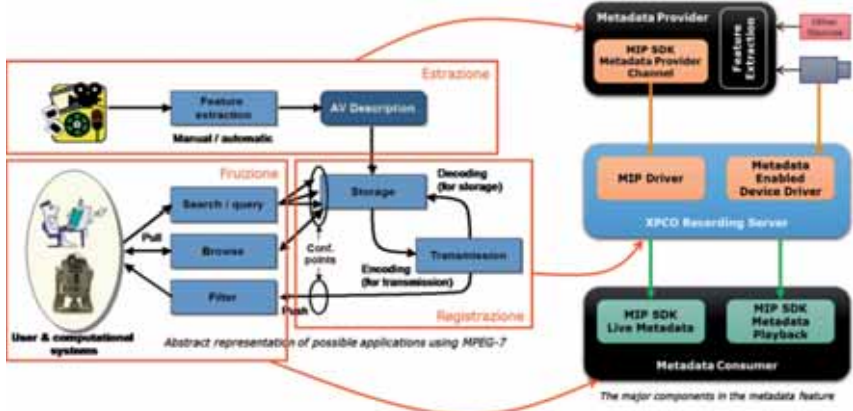
1. Nota storica: la sigla MPEG-3 non fu utilizzata per evitare possibili confusioni con mp3.

2. Nota storica: la sigla MPEG-7 fu derivata proprio dall'applicabilità di tale standard a tutti i 3 precedenti (4+2+1 = 7).



Uno dei produttori che per primo ha implementato sulla propria piattaforma le funzionalità dei metadata, così come specificate da ONVIF, è stato **Milestone** che le ha rese disponibili lo scorso anno con il SDK 2014; nella documentazione è stata utilizzata l'immagine a sinistra per illustrarne i principi di funzionamento.

Da questa ulteriore immagine è facile dedurre che i due schemi in linea di principio coincidano perfettamente:



## ONVIF

Le entità coinvolte nella gestione dei metadata per **ONVIF** sono le seguenti.

Le specifiche dei componenti sono descritte nel documento **Analytics Service Specification**, attualmente rev. 2.5, mentre quelle del WEB Service nel documento "Analytics Service WSDL" (*Web Service Description Language*), attualmente rev. 2.2.

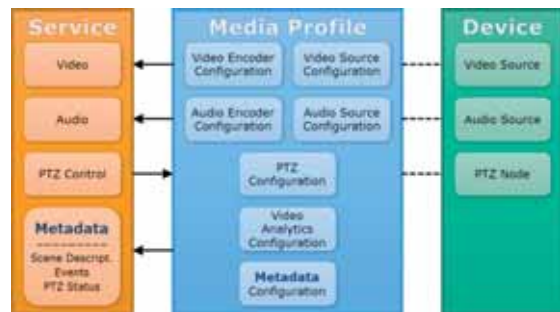
Il paragrafo 5.1 delle Specifiche, intitolato **Scene Description Interface**, elenca gli elementi normati da ONVIF che, alla revisione attuale sono:

**Data/Ora** (*timestamp*) dell'evento

**Posizione** (all'interno della scena)

**Elementi della Scena**, costituiti da:

- **Oggetti**
- **Alberi di Oggetti** (Object Tree), utilizzati per descrivere Oggetti multipli, es. quando due oggetti si avvicinano tanto da non essere più tracciati singolarmente
- **Descrittori delle Sagome** (*Shape Descriptor*), insiemi di elementi geometrici che descrivono il contorno degli oggetti. Al minimo, un descrittore deve contenere il rettangolo che circonda l'oggetto (*Bounding Box*) e il Centro di Gravità (punto che descrive la traiettoria dell'oggetto)
- **Colore**
- **Descrittori delle Celle di Rilevamento degli Oggetti** (*Motion In Cells Descriptor*), la zona della scena dove avviene il rilevamento dell'oggetto
- **Descrittori della Classe degli Oggetti**, che possono essere: **Animali** · **Volti** · **Persone** · **Veicoli** · **Targhe** · **Gruppi** · **Altro**.



È doveroso sottolineare come tali elementi costituiscano il solo insieme di base: è infatti possibile, grazie all'uso del XML (*eXtensible Markup Language*) per la descrizione della scena, aggiungere ulteriori elementi (*tag extension*), oltre a quelli previsti, tramite il costruttore "**Extension**".

La stessa Milestone ha introdotto la prima estensione nel proprio SDK, rendendo disponibile una prima versione (1.0) dei dati di geolocalizzazione, definiti di "Dati di Navigazione", la cui struttura XML è la seguente.

```

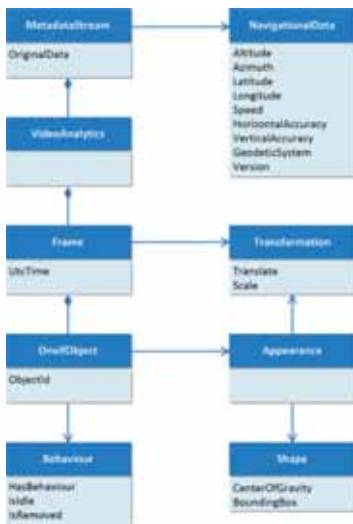
<?xml version="1.0" encoding="UTF-8"?>
<tt:MetadataStream xmlns:tt="http://www.onvif.org/ver10/schema">
  <tt:Extension>
    <tt:NavigationalData version="1.0">
      <tt:Latitude>52.069926</tt:Latitude>
      <tt:Longitude>11.796875</tt:Longitude>
      <tt:Altitude>45.6</tt:Altitude>
      <tt:Azimuth>152.0</tt:Azimuth>
      <tt:HorizontalAccuracy>6.5</tt:HorizontalAccuracy>
      <tt:VerticalAccuracy>6.5</tt:VerticalAccuracy>
      <tt:Speed>36</tt:Speed>
      <tt:GeodeticSystem>WGS84</tt:GeodeticSystem>
    </tt:NavigationalData>
  </tt:Extension>
</tt:MetadataStream>

```

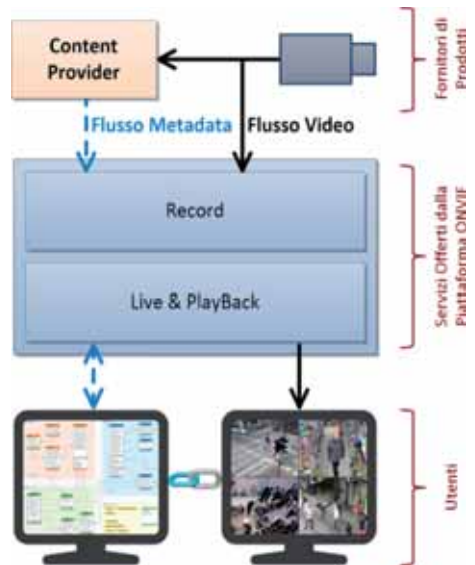
Le definizioni dei campi costituenti il blocco dell'estensione sono quelle riportate in tabella.

Table 1: Definition of fields in Milestone navigation data format	
Field	Description
<b>NavigationalData</b>	This is the container tag for all the navigational data. It is located directly inside the MetadataStream node.
<b>Latitude</b>	In degrees. A double from -90 to +90
<b>Longitude</b>	In degrees. A double from -180 to +180
<b>Altitude</b>	Measured in meters. A double
<b>Azimuth</b>	Aka bearing or course, this is the device angle to true North. Measured in degrees with a double from -180 to +180
<b>HorizontalAccuracy</b>	Horizontal accuracy measured in meters. A positive double
<b>VerticalAccuracy</b>	Vertical accuracy measured in meters. A positive double
<b>Speed</b>	The speed of the device. A non-negative double.
<b>GeodeticSystem</b>	Defines how to interpret the coordinates and altitude. If not present, a value of WGS84 is assumed.

Infine di seguito lo schema delle (principali) classi utilizzate.



È facile immaginare, per esempio, che ai Veicoli possano essere associati altri dati quali la Classe (moto, auto, autocarro, ecc.), il Colore, la Velocità ecc., così come ai Volti il Sesso, la Classe d'Età (giovane, adulto), le Caratteristiche Morfologiche, le Espressioni, ecc. In pratica le informazioni supplementari (Metadata) che possono essere abbinare a un filmato, sono virtualmente infinite. L'immagine a destra illustra il flusso operativo, dalla generazione delle immagini fino alla fruizione dei filmati e relativi metadata, con l'indicazione degli attori in gioco.

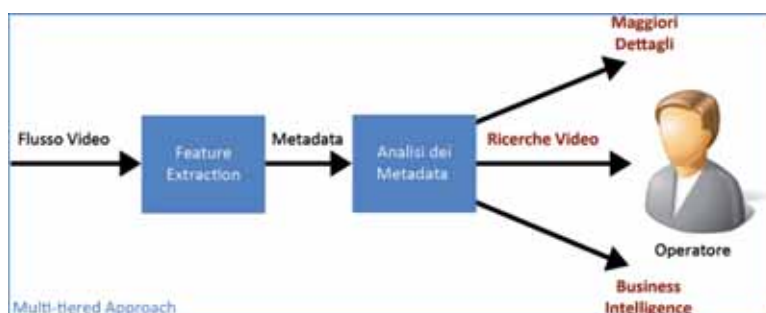


### Potenzialità dei Metadata

In una recente scheda informativa, **Agent Vi** ([www.agentvi.com](http://www.agentvi.com)) utilizza i seguenti schemi per differenziare l'ap-



proccio tradizionale della Video Content Analysis, che definisce Monolitico, da quello innovativo con l'uso dei Metadata, che definisce Multilivello.



E correda tali immagini con la seguente tabella.

	APPROCCIO MULTILIVELLO	APPROCCIO MONOLITICO
<b>Flessibilità</b>	<b>Alto</b>	<b>Basso</b>
	Metadata molto dettagliati forniscono le basi per una varietà di applicazioni	Gli algoritmi dedicati sono ottimizzati per uno scopo specifico. Ulteriori funzionalità richiedono nuovi algoritmi
<b>Esecuzione simultanea di più applicazioni</b>	<b>Facile</b>	<b>Difficile</b>
	I Metadata sono creati una sola volta e utilizzati per varie applicazioni, con il minimo overhead aggiuntivo per ogni ulteriore applicazione	Ogni nuova funzionalità richiede uno specifico algoritmo che incrementa in modo significativo i requisiti di CPU e di memoria
<b>Mix tra real-time e off-line</b>	<b>Facile</b>	<b>Difficile</b>
	La creazione dei Metadata avviene in tempo reale, l'analisi degli stessi può essere intrapresa sia in tempo reale, sia in tempi successivi non in linea	L'intero algoritmo viene eseguito ogni volta sul flusso video, sia esso dal vivo, sia registrato

In sintesi è possibile affermare che l'utilizzo dei Metadata consente di ottenere le seguenti funzionalità aggiuntive:

- Disporre di maggiori **Dettagli degli Eventi** rilevati dal Sistema
- Velocizzare le **Ricerche Video** (degli Eventi) ed effettuare **Correlazioni** tra gli stessi
- Disporre di dati sui quali operare con algoritmi di **Business Intelligence**

### Alcuni Possibili Esempi

#### Dettagli degli Eventi

Supponiamo di disporre di un filtro di analisi che rileva "Accesso Vietato in Zona a Senso Unico", all'accadere dell'evento avremo:

- con sistemi di Video-Analisi tradizionali:
  - notifica dell'evento e relativa osservazione visiva, in alcuni casi distinguendo tra persone e automezzi
- con sistemi corredati da Metadata:
  - notifica dell'evento e relativa osservazione visiva ...
  - Tipo: Veicolo
  - Classe: Autocarro
  - Colore: Rosso
  - Targa: ED126YT
  - Velocità: 45 Km/h

### Ricerche e Correlazioni

Supponiamo di dover verificare se un furgoncino rosso ha attraversato la precedente zona a senso unico:

con sistemi di Video-Analisi tradizionali:

- selezione di tutti gli eventi di attraversamento della zona e successiva osservazione visiva degli stessi alla ricerca del furgoncino rosso

con sistemi corredati da Metadata:

- selezione degli eventi di attraversamento zona con **Tipo = Veicolo**, **Classe = Autocarro** e **Colore = Rosso** e successiva eventuale osservazione visiva dei risultati

Supponendo che la nostra zona a senso unico sia nei pressi di una banca che ha subito un furto, grazie ai dati di cui disponiamo è facile verificare se la presenza del solito furgoncino rosso sia avvenuta, o si sia intensificata nel periodo del furto. Analogamente sarà possibile effettuare ricerche incrociate al fine di verificare se in concomitanza di diversi furti è presente la ricorrenza di una stessa vettura (stessa targa).

### Business Intelligence

Supponiamo di avere una o più telecamere all'interno di un negozio, in corrispondenza di una vetrina, corredate da un sistema di analisi che rileva i volti dei passanti quando entrano nella scena e quando ne escono. L'algoritmo di "Face Detection" è regolato per il rilevamento di volti frontali di una data dimensione minima, per fare in modo che siano esclusi quelli lontani; è plausibile supporre che le riprese siano effettuate su passanti che osservano la vetrina:

- con sistemi di Video-Analisi tradizionali:
  - è *improbabile che, al momento, esistano sistemi di face detection, se non dedicati, che notificano l'uscita di scena*
  - con sistemi corredati da Metadata:
- con i dati a disposizione è possibile ricavare il **Tempo Medio di Osservazione**, per esempio, per **Classe d'Età** (Giovane, Adulto) e **Sesso**, fornendo informazioni su una sorta di "Indice di Gradimento" dell'allestimento della vetrina in funzione di tali categorie
- ma è anche possibile, con algoritmi appena un po' più sofisticati, raccogliere informazioni statistiche sulle **Emozioni** suscitate dalla vista della vetrina, disponendo in tal modo di dati utilizzabili per rendere più attraenti (*emozionanti*) i successivi allestimenti.

In sintesi, è possibile affermare che, tramite specifiche implementazioni SW che utilizzano i Metadata accumulati in un DB, sia possibile impiegarli a un livello superiore a quello della singola Scena, rilevando informazioni più articolate, con maggiore precisione e tempi decisamente inferiori, rispetto a quanto sia possibile fare con le tecniche tradizionali. Sofisticati algoritmi statistici possono consentire l'identificazione di modelli di comportamento dei fenomeni osservati, e fornirne accurate stime quantitative. L'utilizzo di tecniche di **Business Intelligence** sul tali analisi comportamentali permette di stimarne gli "Indicatori", la cui costante osservazione degli scostamenti consente di predire l'occorrenza degli eventi controllati; in pratica...



**Riproduzione riservata**