

# Multimediatatan kuvaus, organisointi ja tallennus

Jussi Hanhijärvi

Jussi.hanhijarvi@tietoenator.com

## Tivistelmä

*Julkaisussa tarkastellaan keskeisiä periaatteita audiovisuaalisen datan kuvaamisesta sellaisessa muodossa, että sitä voidaan joustavasti hakea sisällön mukaisin kuvauksin tietokannoista ja massamuistitallenteista. Kuvaukset voivat olla sekä metadataa, eli tekstuaalista kuvausta sisällöstä että medialähtöistä dataa joka pohjautuu mediasta irroitettuihin piirteisiin, kuten tekstuuriin, muotoon, liikkeeseen tai ääneen. Tarkastelussa keskitytään pääasiassa MPEG-7 standardiin, sen keskeisen merkityksen vuoksi ja siksi että standardi on universaalisen luonteensa vuoksi luonnut joukon sovelluskohtaisia suosituksia.*

Median sisältö on arvokasta vain jos sitä pystytään käyttämään ja sitä voidaan osoittaa luontevasti. Sisältö, johon ei ole helppoa pääsyä voidaan katsoa olevan lähes olemattoman kaikille muille paitsi ehkä luojalleen. Toistaiseksi sisällön hakua on tehty usein selainten avulla käyttämällä sopivia avainsanoja ja indeksejä. Tällainen erilaisilla hakukoneilla tehtävä haku valitettavasti luo huomattavan määrän kohinaa, toiseksi hakuprosessi on kohtalaisen kallista ja hinta kasvaa sisällön määrän kasvaessa, kolmanneksi tekstuaaliset kuvaukset ovat usein sekä subjektiivisia että puutteellisia ja rajoittuneet valitettavan usein vain tiettyyn sovellusalueeseen tai eivät kuvaa median sisäisiä suhteita kovin hyvin. On siis syntynyt kokonaan uusi ongelma-alue: Miten voidaan kuvata, indeksoida ja annotoida kompressoimaton tai kompressoitu audiovisuaalinen data käyttäen työkaluja, jotka automaattisesti irrottavat audiovisuaaliset piirteet sisällöstä ja täydentävät tosinaan käsin tehtyjä tekstuaalisia kuvauksia. Esimerkiksi tietty kuva pitäisi voida löytää piirtämällä kopvaruudulle jonkinlainen muutamalla viivalla tehty luonnos haettavasta, viheltämällä sopiva melodian pätkä pitäisi automaattisten mekanismien pystyä löytämään haluttu musiikkikappale tietovarastosta.

Julkaisu on järjestetty siten, että aluksi esitellään muutama keskeinen ongelma, jotka johti MPEG-7 standardiin. Luvussa 2 käsitellään yleisesti MPEG-7:ää ja tähän läheisesti liittyviä muita standardeja. Lukijan motivoimiseksi esitetään luvussa 3 esimerkkinä still kuva. Tämän jälkeen hahmotetaan yleinen järjestelmä, jolla on mahdollista kuvata audiovisuaalista dataa ns. deskriptoreiden avulla. Tarkastelu on olennaisesti käsitelmällin

selvittämistä. Kuten tuonnempana tullaan huomaamaan niin nämä deskriptorit ovat keskeinen elementti multimediadatan kuvauksessa. Järjestelmäkuvauksen jälkeen kuvataan eräs XML:ään pohjautuva yleinen tyyppimäärittäminen ns. Description Definition Language (DLL) ja siihen tehtävät AV datan erityispiirteet huomioivat laajennukset. Kun aikaisemmat luvut ovat olleet johdantona siihen miten deskriptorita hyödynnetään ja muodostetaan niin luvussa 4 käsitellään varsinaisesti deskriptoreiden sisältöä. Monista tärkeistä seikoista käsitellään luvussa 5 vain perusrakenteita. Hyvin monet aivan keskeiset asiat jätetään lukijan omaehtoisen tutustumisen varaan. Lopuksi luodaan katsaus muutamaan audiovisuaaliseen deskriptoriin. Visualiselta puolelta tarkastellaan värin, muodon ja tekstuurien kuvausta. Audio seikoista tarkastellaan eri keinoja kuvata ääntä. Vaikka näihin alatasoon piirteisiin on kiinnitetty paljon huomiota niin on syytä pitää mielessä, että MPEG-7 on olennaisesti paljon muutakin: metadatasta – hahmontunnistukseen.

## 1 JOHDANTO JA ONGELMAN ASETTELU

80-luvun alusta lähtien on esiintynyt voimakas trendi saattaa audiovisuaalinen data sellaiseen muotoon, että sitä voidaan tehokkaasti tallentaa, hakea ja jakaa digitaalisesti. Luonnollisia näkymiä sisältävälle videolle on syntynyt joukko standardeja, joista uusimmalla H.264:llä voidaan kompressoida PAL laatuista videota lähes 1.2 Mb/s. Samoin on kehitetty joukko audiota kompressoivia algoritmeja joista uusimpana on AAC ja tämän johdannaiset. Keinotekoista AV-dataa varten on myös kehitetty suuri joukko parametrisoituja kuvauksia. Esiintyy myös lukuisia joukko erilaisia yhteisesti sovittuja kuvauksia kokonaisten virtuaalisten maailmojen kuvaamiseksi mainittakoon tässä yhteydessä VRML 97 ja X3D. – Digitaalisessa muodossa esitetyn informaation tuottaminen on siis entistä yksinkertaisempaa. Informaation tallennus on myös halventunut ratkaisevasti. Kun vuonna 2000 1 G kovalevykapasiteettia maksoi n. 10 € niin vuonna 2005 hinnan arvioidaan putoavan n. 1 €/1G tasolle. Käytännössä tämä merkitsee suurentuneen informaation määrän tallentamista myös paikallisesti. Laajakaistaisten yhteyksien halpenemisen ja asennettujen installaatioiden lukumäärän kasvun myötä myös jokaisella on mahdollisuus sekä jakaa että kuluttaa mediaa aivan toisessa mittakaavassa kuin koskaan aikaisemmin.

Koska digitaalisen sisällön luominen on olennaisesti helpottunut niin sen määrä on myös kasvanut räjähdysmäisesti. Jonkinlaisen kuvan kuinka paljon informaatiota tuotetaan maailmassa vuosittain voidaan saada Lyman ja Varianin tutkimuksesta ”How much Information” [LYMA00]. Yhteenvetona he arvioivat, että maailmassa tuotetaan vuosittain 1 – 2 exatavua (miljardi kertaa miljardi eli  $10^{18}$  tavua) informaatiota, vastaten 250 Mtavua jokaista maailman ihmistä kohden. Tästä määrästä tekstimuotoiset dokumentit kattavat vain 0.003 % osuuden lopun jakautuessa optiseen (filmi ja CD/DVD) sekä magneettisen median kesken. Oheiseen taulukkoon 1 on koottu yhteen yksityisten ihmisten luoman informaation määrä (kohtuullisella kompressiosuhteella) [LYMA00]:

Taulukko 1: Yksityisen informaation vuotuinen tuotanto		
	Määrä	Teratavuja
Valokuvia	80 miljardia kuvaa	410 000
Kotivideoita	1,4 miljardia nauhaa	300 000
Röntgenkuvia	2 miljardia kuvaa	17 200
Kovalevyjä	200 miljoonaa installaatiota	13 760
<b>Yhteensä:</b>		<b>740 960</b>

Taulukon tietoja voidaan verrata julkaistun informaation määrään (taulukko 2). Huomataan, että vuonna 2000 yksityinen informaatio oli olennaisesti suurempi kuin julkaistuihin nimikkeisiin sisältyvä määrä. Kummassakin taulukossa on kokonaan jätetty puhelimen, webin, news ryhmien ja s-postiin sisältyvä data koska näitä on, tutkijoiden mukaan, hieman vaikea arvioida :

Taulukko 2: Vuosittain tuotettava julkinen informaatio		
	Nimikkeitä	Teratavuja
Kirjoja	968 735	8
Sanomalehtiä	22 643	25
Aikakausilehtiä	40 000	2
Viikkolehtiä	80 000	10
Tiedotteita	40 000	0.2
Toimistodokumentteja	7 500 000 000	195
Elokuvia	4 000	16
Musiikki CD:tä	90 000	6
Data CD:tä	1 000	3
DVD videoita	5 000	22
<b>Yhteensä:</b>		<b>285</b>

Tämän valtisan informaation hallitsemiseksi olisi toivottavaa, että esiintyisi yleisesti hyväksytyjä mekanismeja, joilla voitaisiin helposti löytää ja yhdistää sisältökomponentteja suuren massan joukosta.

Uusi ongelma-alue on monet agenttipohjaiset järjestelmät. Tässä ajatellaan paradigmaa, jossa älykäs ohjelmisto oppii käyttäjältä hänen kiinnostuksen kohteensa ja käytöstapansa ja pystyy välittämään nämä seikat muille medialäheisille agenteille. Jotta tällainen agentti voisi toimia digitaalista AV dataa sisältävissä ympäristöissä pitää audiovisuaalinen data kuvata sellaisella tavalla, joka luonnehtisi mahdollisimman hyvin datan sisältöä. Toiseksi data pitää voida tallentaa hallitusti ja tavalla joista agenttien välisen vuorovaikutuksen kautta voisi välittyä tieto loppukäyttäjälle.

Kun julkaistun median haku ja paikallinen tallentaminen voidaan tehdä helpoksi, nousee entistä polttavamaksi ongelmaksi tekijöiden immateriaalisten oikeuksien turvaamiseksi. Median kuluttamiseen ja hakemiseen pitäisi olla sisäänrakennettuja mekanismeja, joilla voidaan taata oikeuksien säilyminen ja estää laiton käyttö.

## 2 STANDARDOINTI

Edellä mainittujen ongelmien ratkaisuksi päätti standardointiorganisaatioiden ISO ja ITU puiteissa toimiva MPEG-yhteisö (Motion Picture Expert Group) 1996 Tampereella pidetyssä kokouksessa käynnistää uuden standardointiaktiviteetin, jonka tuloksena luotiin multimedian sisältöä käsittelevä uusi standardi MPEG-7, *Multimedia Content Description Interface* (ISO/IEC 15938). Standardointiaktiviteetti oli valtava, yhtenä tunnusmerkkinä oli se, että ehdotuksia saapui 625, joiden arvioinnin, validoinnin, tuloksena syntyi sarja standardijulkaisuja. MPEG-7 on luonteeltaan ns. toolbox standardi, joka sisältää hyvin suuren mutta kuitenkin minimaalisen määrän erilaisia (toistensa poissulkevia) työkaluja ja menetelmiä (Pereira 2002).

MPEG-7 on luonnollinen jatkumo MPEG yhteisön aikaisemmista standardeista, joita olivat massamuistisovelluksiin tarkoitettu videokuvan kompressiostandardi MPEG-1 ja tämän laajenuksena ensisijaisesti digitaalisiin televisiosovelluksiin tarkoitettu MPEG-2. MPEG-4 standardi taas on suunnattu erityisesti kaiken sekä luonnollisen että keinotekoisien multimedian sisällön kuvaukseen, jakeluun, kompressioon ja keskinäiseen synkronointiin (eikä pelkästään videon kompressioon kuten yleensä luullaan). Perheen uusin tulokas, MPEG-7, poikkeaa aikaisemmista ratkaisevalla tavalla: tässä ei sinänsä käsitellä mekanismeja miten tietty audiovisuaalinen data sinänsä pitää käsitellä. Vaan käsittelee erillaisia tapoja siitä miten AV data voidaan kuvata sekä jakelua että tallennus tapoja varten. Standardi muodostuu kahdeksasta osasta, nimittäin:

- Part 1 Systems:* Määrittelee kaikki ne mekanismit ja työkalut, joita tarvitaan MPEG-7 kuvausten ja työkalujen tehokkaaseen siirtämiseen ja tallentamiseen, sekä sisällön synkronointiin
- Part 2 Description Definition Language:* Määrittelee kielen, jolla voidaan kuvata deskriptorimalleja<sup>1</sup>.
- Part 3 Visual:* Määrittelee deskriptorit ja deskriptorimallit, joita sovelletaan visuaaliseen tietoon.
- Part 4 Audio:* Vastaavasti määrittelee audio datan käsittelyyn soveltuvat deskriptorit ja deskriptorimallit
- Part 5 Generic entities and Multimedia Description Schemes (MDS):* Määrittelee geneeriset, ei siis audio tai video spesifejä, deskriptoreita ja deskriptorimalleja.
- Part 6 Reference Software:* Kuvailee MPEG-7:ään sisältyvien menetelmien ns. referenssi implementaatiota, vertailuihin käytettävää toteutusta.
- Part 7 Conformance Testing:* Määrittelee yhteensopivuustestausprofiilit eli menettelytavat joilla voidaan testata tietyn implementaation standardin mukaisuus.
- Part 8 Extraction and Use of MPEG-7 Description* Julkaisu ei oikeastaan ole standardin normatiivinen osa. Siinä annetaan kuva miten audiovisuaalista dataa voidaan irrottaa osaksi kuvauksia ja miten käytännössä käytetään eräitä deskriptori menetelmiä.

Käytännössä MPEG-7 on osoittautunut olevan tehokas ja robusti standardi. Audio ja visual-osat käsittelevät yksinomaan äänen tai videon kuvaukseen tarvittavia seikkoja, sisältöä kuvaavia piirteitä, kun taas MDS (eli osa 5) esittelee kuvausmalleja, joita tarvitaan AV materiaalin kuvauksessa. Osan 2 DDL:ään sisältyy XML schemaan pohjautuvan määrittelykielen syntaksi ja semantiikka (joista voidaan kehittää myös oma spesifiin tarkoitukseensa soveltuva ”murre”). Osan 1, järjestelmä, tarjoaa taasen välttämättömät vihjeet, jotta standardia voitaisiin soveltaa käytännön tilanteisiin. Lopulta osaan 6 sisältyvä ”open source” referenssi ohjelmisto muodostaa MPEG-7:n erään käyttökelpoisen implementaation.

## 2.1 Esimerkki

Lukijan motivoimiseksi siihen miten MPEG-7:ssä kuvataan mediaa tarkastellaan oheista kuvaa 1 (Anni ja Mikko keskustelevat). Tässä media on still kuva<sup>2</sup>, jossa on luonnehdittu keskeisiä kuvauksessa esiintyviä tekijöitä. Näitä tekijöitä ovat sisällön hallinnointiin, spatiaalisiin ja sisältöön liittyviä seikkoja.

Sisällön hallinnointiin liittyviä tekijöitä ovat luontia, media ja käyttöä kuvailevat tiedot. Luontia luonnehtivilla tiedoilla pyritään vastaamaan kysymykseen kuka, koska ja missä kuva luotiin. Samoin tässä voidaan vastata kysymykseen mitä välineitä käytettiin. Mediaan liittyvillä tiedoilla pyritään hahmottamaan mikä on median formaatti (720 x 437, JPEG koodaus) ja minne se on tallennettu. Käyttötietoihin on usein tallennettu tieto kuka ja millä ehdoilla saa käyttää mediaa. Tämän kaltaiset tiedot liittyvät läheisesti kuvaan mutta eivät varsinaisesti esitä kuvan sisällöstä mitään. Esimerkkikuvan tapaus on sinänsä

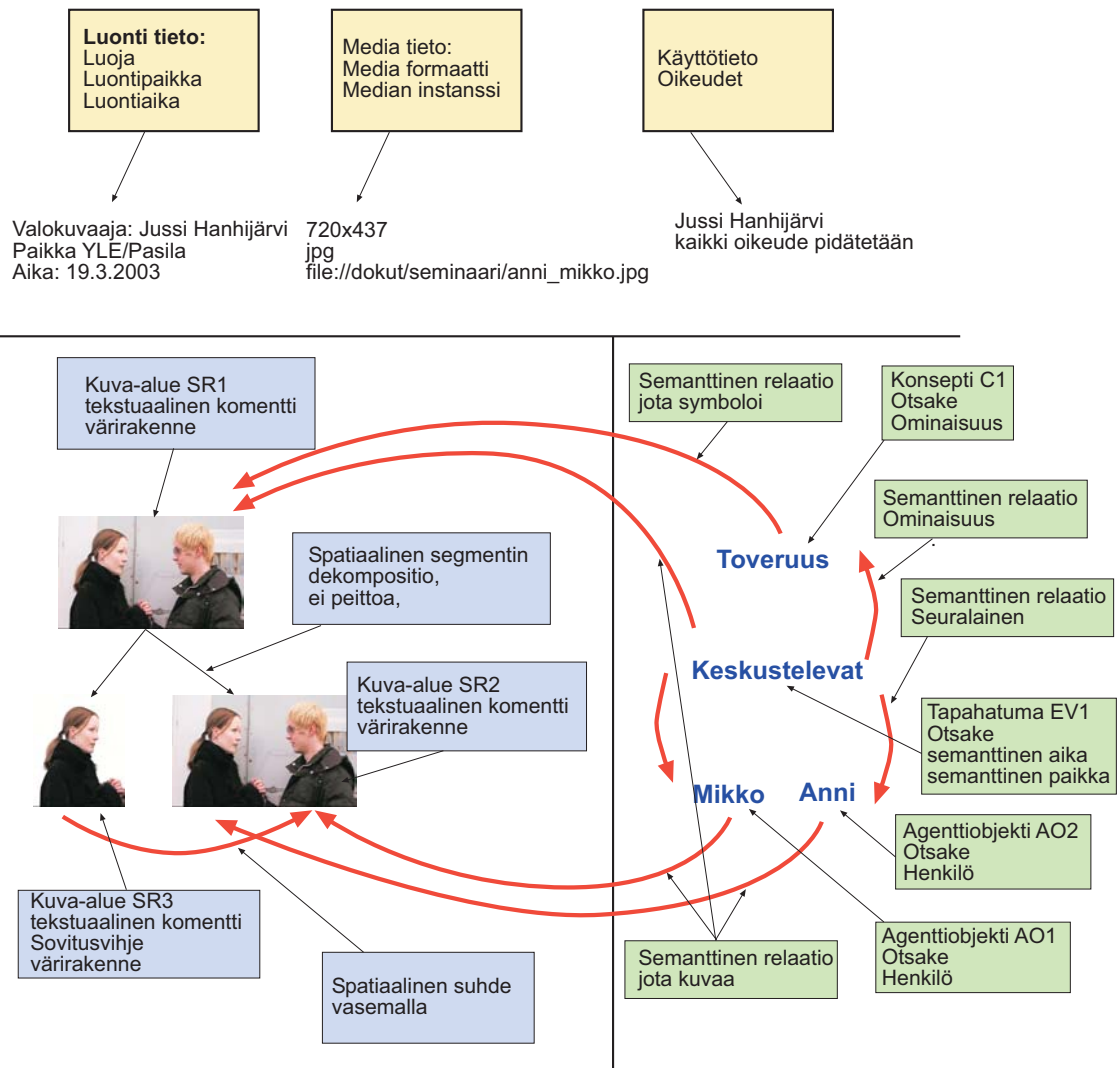
<sup>1</sup> Englanninkieliset vastintermit ovat hieman ongelmalliset. Tässä käytetyn deskriptor vastine descriptor on WSOY:n sähköisen Englanti-Suomi sanakirjan mukaan määrittelijä, kuvaaja, MOT sanakirjaston taasen suosittelee kuvaaja, asiasana, indeksitermi, kuvailutermin vastineita. Tässä on kuitenkin käytetty deskriptor termiä koska kyse on aivan erityisluonteinen median sisällön määrittelijä. Scheme:lle taasen on ehdotettu vastintermejä kaavio, malli jne. Tässä yhteydessä käytetään malli-termiä.

<sup>2</sup> Syynä still kuvan valitsemiseksi esimerkiksi on hyvin yksinkertainen: Paperimuodossa on hieman vaikeata esittää ääntä tai videota

yksinkertainen yksi kuva yhdestä tapahtumasta ja yhdellä koodaustavalla (720x437 JPEG). Mutta usein kuvia voi olla tietystä tapahtumasta paljon, useamman eri kuvaajan kuvaamana, eri kuvakulmilla ja tallennettu eri formaateilla. Tällaista tapahtumaa (esimerkiksi urheilutapahtuma) kutsutaan MPEG-7 kielenkäytössä reality:ksi, jota siis (moninkertainen) media esittää.

Esimerkkikuvasta on koodattu kuvan spatiaalisia suhteita ja alueita kuvaavia seikkoja, merkittynä SR-1, SR-2 ja SR-3. Tässä on esitetty alueiden värirakenne vaikka monia muita seikkoja olisi voitu esittää esimerkiksi tekstuuri ja muoto. Osakuvasta SR-3 on myös esitetty miten se liittyy varsinaiseen kuvaan. Alueiden spatiaalinen dekompositio on myös kuvattu (segmentit eivät peitä toisiaan). MPEG-7:ään sisältyvät sisällön rakennetta esittävät seikat kuvaavat median spatiaalisia ja temporaalisia piirteitä kuten myös lähteen segmenttejä ja näiden välisiä suhteita. Tähän voidaan sisältää hyvinkin mutkikkaita hierarkisia kuvauksia. Esimerkiksi 3-D kuvassa voidaan kuvata osien syvyys suunnassa olevia suhteita, videosegmenteistä voidaan kuvata siirtymiä segmentistä (otoksesta) toiseen, kollaasista osakuvien suhdetta kokonaisuuteen jne. MPEG-7:n kuvaustavat ovat tässä erittäin monipuoliset.

Esimerkin kolmas osa-alue muodostuu kuvan semantiikasta eli siitä mitä kuva oikeastaan esittää: Anni ja Mikko pitävät toisiaan kädestä ja keskustelevat ystävällisesti keskenään. Multimedian kontekstuaalinen (sisällön) semantiikka usein aloitetaan kuvaamalla mediassa esiintyviä tapahtumia (esimerkiksi keskustelevat, pitävät toisiaan kädestä). Kuvaukseen voidaan lisätä ihmiset ja organisaatiot (joita kutsutaan MPEG-7 kielenkäytössä agenteiksi), tapahtumapaikat, rakennukset jne. Lisäksi kaikilla näillä tapahtumaan osallistuvilla seikoilla on keskinäinen suhde (esimerkiksi Anni ja Mikko seisovat oven edessä) ja erilaisia tiloja. Tapahtumat tapahtuvat tiettyyn taustaan jne. MPEG-7 sisältää hyvinkin



monipuoliset työkalut semanttiseen abstraktointiin (esimerkiksi toveruus, kaksi ihmistä vuorovaikuttavat keskenään) joita voidaan tarkentaa erilaisin attribuutein (Anni, Mikko, keskustelevat). Samoin abstraktioiden ja niiden attribuuttien välisiin suhteisiin esiintyy myös laaja valikoima erilaisia menettelyitä.

## 2.2 Muita läheisesti liittyviä standardeja ja ehdotuksia

Vaikka MPEG-7 tarjoaa erinomaiset työkalut mediamaisemassa navigointiin niin se ei kuitenkaan ole ainoa alueella toimiva standardi. Muut lähinnä ”metadataan” orientoituneet (monasti MPEG-7:ää täydentävät) standardit ovat suuntautuneet enemmän tai vähemmän tietyn erityisalueiden tarpeisiin. Oheiseen taulukkoon on kuvattu muutama keskeinen ehdotus ja toetus:

Julkaisu	Käyttötarkoitus
Harmony	Ensisijaisesti digitaalisten kirjastojen käyttöön tarkoitettu multimedian kuvaus. Pyrkii täydentämään sovellusalueellaan MPEG-7:ää ja Dublin Corea
Dublin Core	Kehitetty ensisijaisesti kuvaamaan tekstuaalisessa muodossa esitettyjen kirjastoituja, museoituja tms. materiaalia
ANZLIC (Australia New Zealand Land information Council)	Lähinnä maantieteellisiin sovelluksiin ja rakentamiseen kehitetty spatiaalisen informaation metadatan esitystapa.
GILS (Global Information Locator Service)	Ensisijaisesti geospaatialisen informaation hakemiseen tarkoitettu metadatan kuvaustapa.
EdNA (Educational Network Australia)	EdNA on määritellyt kohtalaisen yksinkertainen e-learning sovelluksiin tarkoitettun metadatan kuvaustavan.
IMS (Instructional Management System)	Muiden asioiden lisäksi IMS on kehittänyt e-learning sovelluksiin tarkoitettun laajennettuun Dublin Coreen perustuvan kuvaustavan.
GEM (Gateway to Educational Material)	Dublin Core:sta laajennettu kuvaustapa e-learning sovelluksiin
vCard (Electronic Business Card)	Toimikortti transaktioihin kehitetty mediaan liittyvän metadatan kuvaus.
A-Core (Metadata about Content Metadata)	IETF (Internet Engineering Task Force) alaisuudessa laadittu internetin välityksellä siirrettävän AV materiaalin metadatan kuvaus.
MXF	Pro-MPEG forumin julkaisema broadcast ympäristöön tarkoitettu tiedostoformaatti.

Näistä ”metadata” standardeista Dublin Core on keskeinen. Nimi juontaa juurensa siitä, että ehdotus laadittiin Dublin:n kaupungissa (Ohio ) vuonna 1995. Alkuperäisestä kirjastoihin tarkoitettua tekstuaalisia dokumentteja käsittelevästä ehdotuksesta on se laajentunut moniin valtionhallinnon ja e-learning dokumenttien hallinta sovelluksiin. Uusin versio sisältää 15 dokumentin hakua helpottavaa elementtiä, jotka kuvaavat esimerkiksi dokumentin nimeä, luoja, kustantajaa, kieltä, aihealuetta jne. Myös ei-tekstuaaliset dokumentit kuten video ja still kuvat sisältyvät uusimpiin Dublin Coren versioihin [HUNT ]. MPEG-7 kuvaukset voidaan konvertoida Dublin Coren käyttämään muotoon, vaikka konversio on kohtalaisen mutkikas ja näin voidaan Dublin Coren katsoa olevan MPEG-7 osajoukko niiltä osin, jotka koskevat mediariippumatonta hakua.

MXF taasen on ns. profesional MPEG forumin laatima tiedostojen kuvaustapa, joka on tarkoitettu broadcast median vaihtoon eri palvelimien, massamuistien ja digitaalisten arkistojen välillä. MXF paketoit kokonaisuuksiksi audion, videon, ohjelmiin liittyvän tekstidatan kuten myös liittyvän metadatan. MXF on ensisijaisesti tiedostoformaatti, joka alun perin syntyi broadcast yhtiöiden ohjelmien vaihtotarpeisiin.

### 3 JÄRJESTELMÄ

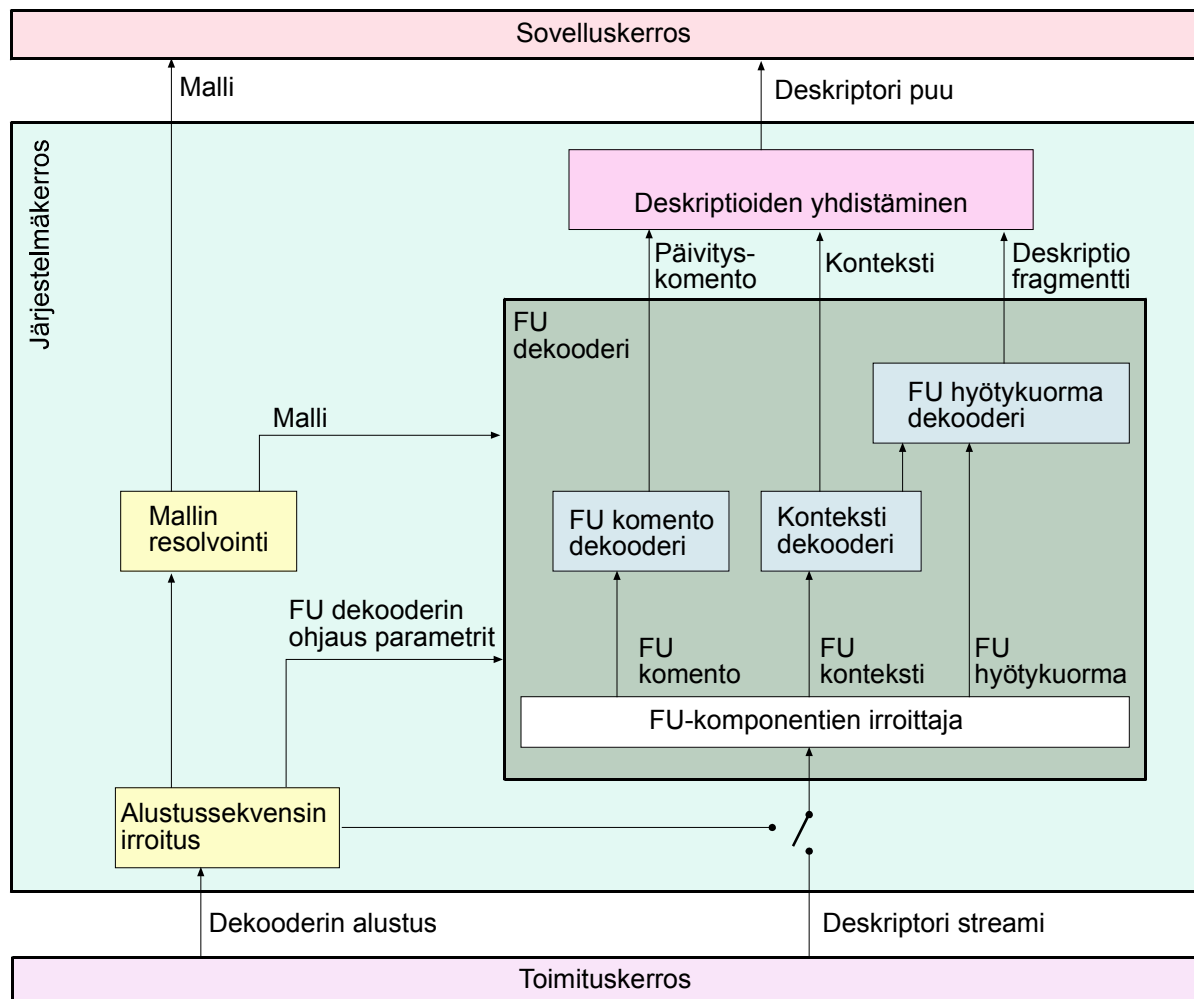
Eräs MPEG-7:n tavoite oli se, että erilaiset kuvaukset seuraavat AV-dokumenttien mukana aina clientille saakka (osa dokumenteista esimerkiksi storyboard voi olla syntynyt jo ennen varsinaista mediataltiota). Verkkoympäristössä clientina ajatellaan median kulutuslaitetta esimerkiksi koti-PC:tä, broadcast ympäristössä taasen clientti on tosinaan lähetyjärjestelmä ja toisinaan kotivastaanotin. Niinpä MPEG-7:lle asetetut vaatimukset voidaan luokitella kahteen luokkaan: Ensinnäkin perinteiset medialähtöiset vaatimukset ja toiseksi spesifit kuvauksiin ja niiden jakeluihin liittyvät vaatimukset (Avaro 2002).

1. *Jakelu.* Multimedia deskriptorin jakelu voidaan tehdä erilaisten protokollien alaisuudessa rinnan varsinaisen median kanssa, esimerkiksi live broadcast voidaan streamata UDP/RTP protokollalla sopivasti valituin pakettirakentein. Monissa menettelyssä voidaan deskriptorit paloitella pienempiin osiin, jotta median kanssa rinnakkainen jakelu kaistarajoitteisissa verkoissa onnistuisi tehokkaasti.
2. *Synkronointi.* Yleensä audiovisuaalinen datan eri komponentit ovat ajallisesti läheisesti kytköksissä keskenään. Monissa tapauksissa deksriptoreihin sisältyvä informaatio pitääkin esittää täsmällisesti määritellyissä ajanhetkissä yhdessä varsinaisen sisällön kanssa.
3. *Streamin hallinnointi.* Koko audiovisuaalisen streamin mukaan lukien deskriptorien hallinnointi vaatii erityisen laatuista hallinnointimekanismeja. Esimerkiksi data voi sijaita hyvin eri paikoissa, datatyyppien tunnistaminen ja niiden keskinäiset suhteet voivat olla hyvinkin vaihtelevia jne.
4. *Esitysmuoto.* Deskriptorit ja deskriptorimallit pitää esittää erityisessä formaatissa täsmällisesti määritellyllä syntaksilla ja semantiikalla. Syntaksien kuvaamiseen MPEG-7:ssä on määritelty ns. deskriptoiden kuvauskieli DLL. DLL pohjautuu XML Schemaan.
5. *Inkrementaalinen lähetys.* Monissa ympäristöissä mihin MPEG-7:ää sovelletaan ovat resurssit (esimerkiksi kovalevykapasiteetti tai verkon kaistaleveys) sangen niukkoja. Tämän vuoksi dataa pitäisi voida lähettää inkrementaalisti tai kompresoidusti. Inkrementaalinen lähetys tehdään ns. Acces Unit:ien (AU) avulla ja vastaavasti kompressoitu lähetys suoritetaan ns. BiM formaatilla (Binary format for MPEG). BiM:n avulla pyritään muodostamaan kompakti ja streamattava esitysmuoto deskriptoreista.

#### 3.1 Päätelaitteen arkkitehtuuri

Oheiseen kuvaan 2 on hahmoteltu tyypillinen MPEG-7:n päätelaitteen arkkitehtuuri. Päätelaitteena ymmärretään tässä olevan joko itsenäinen sovellus tai osa sovellusjärjestelmää. Arkkitehtuuri muodostuu olennaisesti kolmesta kerroksesta. Nimittäin ns. toimituskerroksesta (delivery layer), järjestelmäkerroksesta (system layer) ja varsinaisesta sovelluskerroksesta. MPEG 7 ei varsinaisesti määrittele toimituskerrosta eikä tapaa miten sovellus hyödyntää kuvausta:

- *Sovelluskerros* muodostuu varsinaisesti median käyttöön tai kulutukseen liittyvistä seikoista. MPEG-7 ympäristössä median esittämisessä käytetään siihen liittyviä deskriptoreita ja deskriptorimalleja.
- *Järjestelmäkerros* osana MPEG-7 normatiivista standardia käsittelee menettelyitä miten deskriptorivirrasta (streamista) voidaan muodostaa sovelluskerroksessa hyödynnettäviä deskriptoreita ja deskriptoripuita. Järjestelmäkerros sisältää olennaisesti erilaiset alustustoiminnot, ns. Fragment Unit dekooderin, jolla muodostetaan sovelluksessa käytettäviä valideja deskriptoreita alkeellisempien tekijöiden (FU:den) avulla. Lopulta alkeellisemmat deskriptorit voidaan yhdistää deskriptoripuiksi.
- *Toimituskerros*, osana liikennöintikerroksia, huolehtii virtojen vastaanotosta ja lähetyksestä erilaisten protokollien alaisuudessa.



Kuva 2: MPEG 7:n järjestelmärakenne Avaron & al. mukaan (Avaro 2002).

Varsinaisen toimituskerroksen ulkopuolella esiintyy tallennettu AV-data ja siihen liittyvä infrastruktuurit esimerkiksi tallennettuna kovalevyille tai muulle massamuistilaitteelle. Nämä multipleksoidaan vaikkapa MPEG-2 transport streamiksi, eli paketoidaan 188 tavua pitkiin äänen, videon ja muun datan kesken lomitettuihin paketteihin. Toimituskerros sisältää mekanismit millä voidaan muodostaa deskriptoreiden synkronointi, kehystys ja multipleksointi eli muodostetaan elementääriset kuvausstreamit. Nämä voidaan lähettää joko varsinaisen sisällön yhteydessä tai näistä riippumatta. Kuvausstreamit taasen muodostuvat yhdestä tai useammasta yksilöllisesti osoitettavasta data-alkiosta ns. Access Unit:sta, AU:sta. Jokainen AU jakautuu taasen



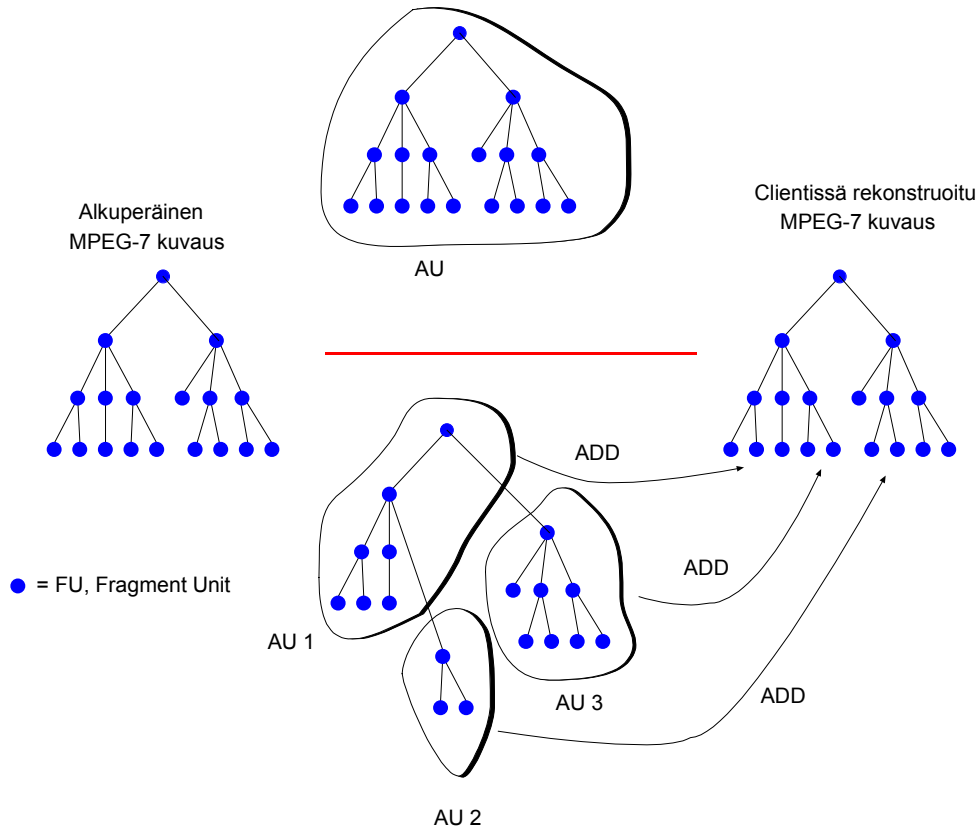
pienempiin osiin ns. Fragment Unit:ksi FU:ksi. AU voidaan luonnehtia siten, että se on pienin datayksikkö mihin voidaan liittää erilaisia ajoitustietoja. Nämä ajoitustiedot taasen ovat suureita jotka kertovat milloin tietty AU ja tähän liittyvä deskriptori (deskriptori malli) aktivoituu. Kuten edellä mainittiin niin AU:t ovat keskeinen elementti inkrementaalisisessa lähetyksessä.

### *3.1.1 Dekooderin alustus*

Kuvassa 1 esiintyi erillisinä suorituspolkuna dekooderin alustus. Tämän tehtävänä on vastaanottaa toimituskerroksesta erilliset alustuskomennot. Dekooderin alustus sisältää luettelon ns. Schema Uniform Resource Identifier:stä (URI). Nämä URI:t määrittelevät ne aktuaaliset käytettävät mallit (schemat), joiden avulla tarkistetaan media kuvauksen (deskriptoreiden) syntaktinen oikeellisuus päätelaitteessa. Lisäksi alustusvaiheessa asetetaan erilaisia parametreja, joilla konfiguroidaan dekooderi. Schema URI:t prosessoidaan ns. Schema Resolverilla, jonka ulostulona ovat deskriptorimallit johdetaan ns. FU dekooderille, joka prosessoi ja kelpoistaa (validoi) deskriptorit. Lopullisena tavoiteena FU dekooderilla on luoda sovellukselle käyttökelpoinen deskriptoripuu. Malli (schema) resolveri ei ole varsinaisesti osa standardia ja se voi noutaa käytettävät mallit erikseen tai käyttää etukäteen tallennettuja malleja. Mikäli aktuaalinen URI tai jokin tämän osa on dekooderille tuntematon niin vastaavat tietotyypit deskriptorivirrassa yksinkertaisesti hylätään. Monissa tapauksissa hylkäämisestä ei ole haittaa, koska hylätty datatyyppi voi olla sellainen mitä tietyn kohteeseen suunniteltu sovelluskerros ei juuri pysty hyödyntämään. Esimerkiksi jos tietyn deskriptori sisältää tietyn äänitteen luontia luonnehtivia suureita (vaikkapa levy-yhtiön tunnisteiden) ja kuluttajalaite ei tarvitse tätä tietoa niin se hylätään.

### *3.1.2 Access Unit rakenne*

Kuavaus virrassa yksilöllisesti osoitettavat AU:t ovat strukturoitu joukoksi komentoiksi, joilla voidaan kapseloida ja ohjata MPEG-7 deskriptoreita.. Komentojen avulla voidaan luoda joustava kuvausten dynamiikka: Deskriptorit ja niihin sisältyvät FU:t voidaan lähettää kokonaisuina ryhmänä tai jaettuna pienempiin osiin ja lähettää inkrementaalisti. Oheiseen kuvaan 2 on hahmoteltu eräs deskriptoreiden jako Access Unitteihin (AU:hin).. Deskriptorit muodostavat puumaisen rakenteen (ns. Deskriptor Tree:n). Puun solmuihin on tallennettu sisällön kuvaukset ja liitynnät muihin solmuihin. Kuvan yläosassa yhteen AU:hun on kapseloitu koko deskriptoripuu kun taasen alasosaan on puu kapseloitu 3:een erillisiin AU:hun. Päätelaitteessa yhdistetään nämä yhdeksi deskriptori puuksi AU:hun sisältyvällä ADD komennolla.



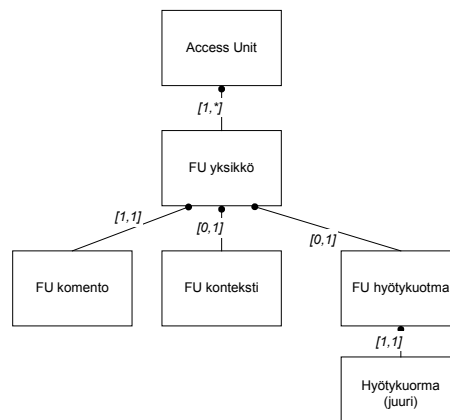
**Kuva 2:** Eräs esimerkki jakaa deskriptorit eri kokoisiksi osoitettavissa oleviksi AU:ksi

### 3.1.3 Fragment Unit rakenne

Kun dekooderi on alustettu niin AU:den prosessointi voi alkaa. Jokainen AU voi muodostua mielivaltaisesta määrästä FU komponenteista (Fragment Unit), joista jokainen irrotetaan itse sekvensistä erikseen prosessointia varten. FU sisältää 3 seikkaa:

- *FU komento* määrittelee suoritettavan päivityksen tyyppin. Mahdollisuuksia on 4. a) lisää fragmentti, jossa lisätään deskriptoripuun solmuun uusi lapsi, b) korvaa fragmentti, c) tuhoa fragmentti (tietystä solmussa ja tämän lapset) d) reset, palauta koko deskriptori alkuarvoonsa.
- *FU konteksti* identifioi käytetyn datatyypin deskriptorimallissa (scheman), sekä osoittaa FU:n sijainnin prosessoitavassa deskriptiopiussa
- *FU hyötykuorma* kuvailee deskriptioon sisältyvän osan varsinaisen arvon, jota edelleen prosessoidaan hyötykuormadekooderilla.

Kun deskriptorit esitetään tekstuaalisesti niin FU hyötykuormadekooderi on itse asiassa parseri, joka jäsentää dokumenttien tyyppimäärytyksiä (DDL:liä). FU komento- ja kontekstimääreillä asetetaan FU-dekooderilta vastaanotettu deskriptorilapsi sopivaan solmuun deksriptoripuussa (sen oikealla käyttöhetkellä). Kun lopulta koko deksriptoripuu on rekonstruoitu FU-dekooderissa, niin voidaan puuta käyttää sovelluserroksessa multimedian kuluttamiseen tietyllä deskriptorissa määritellyllä ajanhetkellä tai aikajaksolla. Oheiseen kuvaan on hahmoteltu UML kaavio AU:n rakenteista FU:den avulla.



**Kuva 3:** Access Unit:n rakenne (Avaro 2002) (hakasulkujen sisälle haarojen minimi ja maksimimäärät)

## 4 KUVAUSKIELI DDL

Eräs keskeisimpiä piirteitä MPEG-7:ssä on käytetty kuvauskieli ns. Description Definiton Language (DDL). Kielellä määritellään multimediadatan liittyvien kuvausten varsinainen sisältö. DDL ei ole mallinnuskieli kuten UML vaan mallikieli, schema-kieli, jolla voidaan kuvata datamalleja. DDL määrittelee deskriptoreiden ja deskriptorimallien syntaksin, rakenteen ja rakenne-elementtien arvot. Hunterin mukaan DDL:n pitää kyetä ilmaisemaan deskriptoreihin sisältyvien elementtien strukturaaliset, periytymis-, spatiaaliset, temporaaliset, spatiotemporaaliset ja käsittelliset piirteet sekä näiden väliset suhteet (Hunter 2002). DDL pohjautuu hyvin pitkälle XML-Schema kieleen, vaikka tämä ei ole suunniteltu AV datan kuvaamiseen. Tämän vuoksi DDL:ään on lisätty multimediakuvausten vaatimat tietorakenteet.

XML schema muodostuu kolmesta schema komponenttien muodostamasta luokasta. Näitä ovat ensisijaiset, toissijaiset ja avustavat komponentit (Hunter 2002).

XML Scheman ensisijaiset komponentit ovat sellaisia, joilla voidaan muodostaa yksinkertaisia XML kuvauksia jo yksinään:

- *Nimiavaruudet* ovat kokoelma nimiä, joita voidaan käyttää dokumentin elementtien ja attribuuttien tunnisteenä. Nimiavaruudet muodostavat yksinkertaisen metodin joilla voidaan muodostaa yksinkertaisia tunnuksia, joita voidaan uudelleen käyttää muissa XML dokumenteissa. Ns. kvalifoidut nimet muodostuvat alkuosasta, kaksoispisteestä ja paikallisesta osasta esimerkiksi `mpeg7:VideoSegmentType`.
- *Elementtienselvityksellä* esitetään elementin tyyppin määrittäminen joko suoraan eksplisiittisesti tai referenssinä tai sillä voidaan kertoa oletusarvon esiintymisen. Esimerkiksi elementtimäärittämisessä nimeen `Country` liitetään `countryCode` oletusarvolla en seuraavasti:

```
<element name="Country" type="countryCode" default="en" minOccurs =
"0" maxoccurs="unbounded" />
```

- *Attribuuttiselvityksillä* mahdollistetaan tiettyjen lisämääreiden esiintymien dokumenteissa liittämällä attribuutti yksinkertaiseen datatyyppiin. Attribuuttiselvityksissä `use` tarkentimella voidaan määrittää attribuutin käyttö ja esiintyminen (`required|optional|prohibited`). Esimerkiksi:

```
<attribute name="lang" type="language" use="optional" default="fi" />
```

```
<element name="Esimerkki">
  <complexType>
    <attribute ref="lang" />
  </complexType>
</element>
```

Esimerkin mukaisesta tyyppistä voidaan muodostaa validi instanssi:

```
<Esimerkki lang="en-us"/>
```

- *Tyypimääritykset* eroavat XML Schemassa ratkaisevasti selvityksistä (declaration). Selvitykset mahdollistavat dokumenttiin sisältyvien attribuuttien ja elementtien esiintymisen määrätyillä nimillä. Tyypimääritykset taas määrittelevät uudet Schema komponentit, joita voidaan edelleen käyttää muissa schema komponenteissa kuten elementtien ja attribuuttien selvityksissä. Tyypimääritykset voidaan jakaa yksinkertainen, kompleksinen, johdettu ja anonyymi luokkiin. Esimerkiksi voimme aluksi määritellä 7 merkkisen merkkijonon "postCode":

```
<simpleType name="postCode">
  <restriction base="string">
    <length="7"/>
  </restriction>
</simpleType>
```

Nyt voimme selvittää elementtejä tai attribuutteja käyttäen tätä uutta "postCode" määritystä. Esimerkiksi voimme määritellä elementin "MyPostCode"

```
<element name="MyPostCode" type="postCode"/>
```

Toissijaisiksi komponenteiksi on määritelty ne komponentit, joilla voidaan luoda ja nimetä erilaisia attribuuttien tai elementtien muodostamia ryhmiä. Näitä voidaan sitten käyttää edelleen monimutkaisemmissa tyypimäärityksissä. Esimerkiksi attribuutit voidaan ryhmitellä:

```
<attributeGroup name="person_attributes">
  <attribute name="weight" type="decimal"/>
  <attribute name="height" type="decimal"/>
  <attribute name="age" type="integer"/>
</attributeGroup>
```

Tämän avulla voidaan määritellä kompleksisen tyyppin "Person" attribuutit

```
<complexType name="Person">
  <sequence>
    <element ref="Name"/>
  </sequence>
  <attributeGroup ref="person_attributes"/>
</complexType>
```

Kolmanteen XML-Scheman luokkaan kuuluvat erilaiset avustavat komponentit. Nämä eivät voi esiintyä XML dokumentissa itsenäisesti. Näitä ovat esimerkiksi erilaiset kommentit, "jokerikortit" (wildcards) jne.

XML-Schema sisältää hyvin laajan kokoelman erilaisia datatyypppejä. Näitä ovat sekä primitiiviset tyypit kuten floating point luku, hexaluvut, kellonaika, kesto että johdetut datatyyppit kuten merkkijono, kielen tunnus, erilaiset kokonaisluvut jne. Johdetuille tyypeille voidaan asettaa erilaisia reunaehtoja ns. facetteja, kuten numeroiden lukumäärä, enumeraatiot, minimi ja maksimiarvot jne. (Hunter 2002).

## 4.1 MPEG-7 laajennukset

Näiden lisäksi MPEG-7 asettaa kaksi laajennusta ensinnäkin taulukko ja matriisidatatyypit ja toiseksi kaksi sisäänrakennettua johdettua datatyyppiä.

MPEG-7 asettaa taulukko ja matriisi DDL mekanismeille vaatimuksena (Hunter 2002):

- Moniulotteiset matriisit pitäisi voida rajoittaa etukäteen määriteltyihin facet-arvoihin mallin (scheman) määrittelyn yhteydessä
- Mahdollisuuden rajoittaa yksiulotteisen taulukon tai moniulotteisen matriisin koon instantiointihetkellä esiintyvään attribuutin mukaan (dynaaminen koon määrittely).

list datatyypin avulla voidaan määritellä yksiulotteisen taulukon ja moniulotteisen taulukon tyyppi ja koko. Tämän ”sisään” määritelty facetti `mpeg7:dimension` muodostuu listasta positiivisia kokonaislukuja, joilla voidaan määritellä kiinteämittaisen matriisin dimensiot. Jotta nämä uudet tyypit menisivät XML parserin ”läpi” joudutaan ne käärimään `<annotation><appinfo>` tagien väliin. Oheinen esimerkki esittää miten voidaan muodostaa ja instantioida kiinteä 3 x 4 kokoinen kokonaislukumatriisi:

```
<simpleType name="IntMatrix2D">
  <list itemType="Integer">
    <annotation><appinfo>
      <mpeg7:dimension value="unbounded unbounded"/>
    </appinfo></annotation>
  </list>
</simpleType>

<simpleType name="IntMatrix3x4">
  <restriction base="IntMatrix2D">
    <annotation><appinfo>
      <mpeg7:dimension value="3 4" />
    </appinfo></annotation>
  </restriction>
</simpleType>
```

Lopuksi matriisi voidaan instantioida tiettyihin arvoihin seuraavasti:

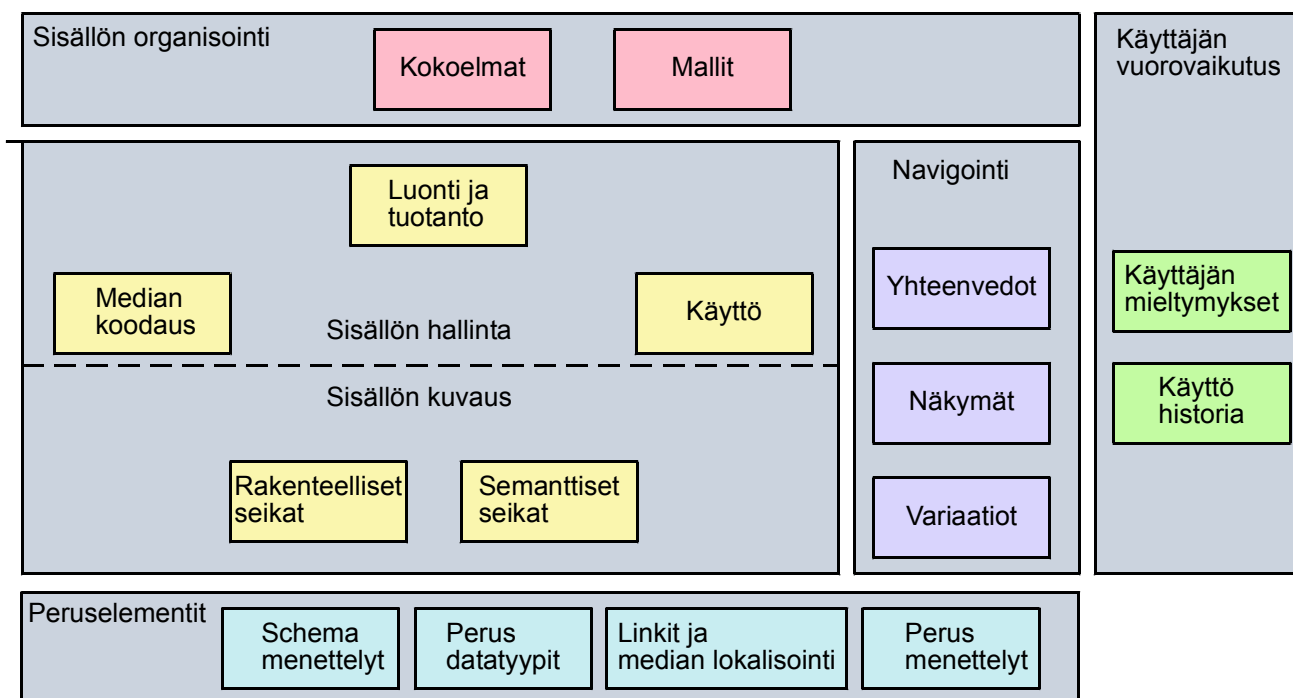
```
<element name="IntMat3x4" type="IntMtrix3x4"/>
<IntMat3x4>
  5 8 9 4
  7 6 1 2
  1 3 5 8
</IntMat3x4>
```

MPEG-7:ään on lisätty kaksi aikasuureita kuvaavaa datatyyppiä. Ensinnäkin `basicTimePoint` tyyppi määrittelee ajanhetken Gregoriaanisina päivinä ja kellonaikoina. Toiseksi `basicDuration` määrittelee aikajakson keston keston päivinä tai päivin osina (h, m,s) sekunnin murto-osiin käytetään `basicTimePoint` datatyyppiä.

## 5 MULTIMEDIA DESKRIPTORI MALLIT (MDS)

Kun MPEG-7:n deskriptorit on suunniteltu kuvaamaan multimedian sisällön yksittäisiä piirteitä niin multimedian deskriptorimalleilla (MDS multimedia description scheme) muodostetaan kompleksisempia kuvauksia yhdistämällä deskriptoreita ja alkeellisempia deskriptorimalleja yhteen.

Salembert & al. Mukaan MPEG-7 deskriptorit on suunniteltu kuvaamaan kahden tyyppistä informaatiota (Salembert 2002): Ensinnäkin medialäheisiä audiovisuaalisia piirteitä (ns. alatason piirteinä), joita ovat esimerkiksi väri, tekstuuri, liike, audion energia jne. Toiseksi abstraktimpia, korkeamman tason, piirteitä kuten tapahtumat, sisällön abstraktit kuvaukset, sisällön hallintaprosessit, kuvaukset tallennusmediasta jne. Alatason medialäheiset piirteitä voidaan usein muodostaa automaattisesti kun taas ylemmän tason piirteet vaativat enemmän tai vähemmän ihmisen toimenpiteitä.



Kuva 4: Yleiskuva MPEG-7 media deskriptorimallien elementeistä

Oheiseen kuvaan 4 on hahmoteltu MDS elementit. Nämä voidaan ryhmitellä 6:een ryhmään, nimittäin peruselementteihin, sisällön hallintaan ja kuvaukseen, sisällön organisointiin, navigointiin ja käyttäjän vuorovaikutukseen. Oheessa esitämme lyhyen ylimalkaisen kuvauksen näistä.

## 5.1 Peruselementit

Peruselementteihin kuuluvat sellaiset osat joita käytetään toistuvasti läpi koko deskriptori mallien määrittelyssä (Walker 2002). Näitä ovat mallien (schemojen) muodostus menettelyt, perusdatatyytit, linkit ja median lokalisointi (vrt myös lukuun 6).

Schema menettelyillä tarkoitetaan tässä yhteydessä niitä tapoja, joilla voidaan muodostaa valideja MPEG-7 deskriptoreita ja paketoita. Paketoinnilla organisoidaan ja helpotetaan sisällön navigointia. Schema menettelyihin sisältyy myös erilaiset metadatat itse deskriptorista. Näitä ovat esimerkiksi erilaiset tunnistetiedot kuten versio, julkisuusaste jne.

Perusdatatyypeihin on kuten edellä mainittiin DDL: n yhteydessä sisällytetään mm. erilaisia ajoitustietoja. Samoin niihin on sisällytetty audion ja videon deskriptorit. Tunnempänä käsittelemme yksityiskohtaisemmin varsinaisia deskriptoreihin sisältyviä perusdatatyyppiä ja niiden soveltamista audion ja videon kuvaukseen.

## 5.2 Sisällön hallinta

Keskeisiä kysymyksiä sisällön hallinnasta ovat mm. kuvaukset luonnista, tavasta koodata media ja kysymykset siitä miten mediaa käytetään (Benitez 2002):

- Informaation luontia kuvaavissa tekijöissä kuvataan miten multimedian sisältö on tuotettu. Tähän kuuluu mm otsake (joka voi olla tekstuaalinen tai toinen multimediaa

sisältävä dokumentti), kuvaus tekijöistä, luontipaikasta ja päivistä. Lisäksi tähän voi sisältyä erilaisia luokittelutietoja kuten genre, esiintyvät subjektit, käyttötarkoitus, kieli jne.

- Tiedon käyttöön liittyvissä kuvauksissa määritellään mm. käyttöoikeudet, käytettävät tietueet sekä erilaisia taloudellisia seikkoja luonnehtivat seikat.
- Median koodausta kuvaavissa elementeissä kerrotaan mm. kompressiotapa ja sisällön tallennusformaatti. Samoin siihen sisältyy tieto ns. mastermediasta eli alkuperäisestä lähteestä, josta on muodostettu erilaisia sisällön ilmenemismuotoja (esimerkiksi eri enkoodaus ja tallennustavoilla tehtyjä ilmentymiä).

### 5.3 Sisällön kuvaus

MPEG-7:ään sisältyy myös sisällön kuvaukset. Näissä kuvataan rakenne (esimerkiksi videon sisältyvät framit ja audion segmentit) ja sisällön semantiikka (objektit, tapahtumat ja erilaiset abstraktit notaatiot) (Benitez 2002):

- Sisällön kuvauksen strukturaalisilla seikoilla kuvataan multimedian sisältö rakenteelliselta katsantokannalta nähtynä. Tässä käytetään segmentoituja kuvauksia jotka kuvaavat spatiaalisia, temporaalisia tai spatiotemporaalisia piirteitä multimedian sisällöstä. Segmentoidut kuvaukset voidaan järjestää hierarkisesti indeksoidun haun helpottamiseksi. Segmentit voidaan kuvata myös havaittavien piirteiden mukaan esimerkiksi värin, tekstuurin, muodon, liikkeen jne. piirteiden mukaisesti järjestettynä.
- Sisällön kuvauksen käsitteellisissä seikoissa taasen katsantokanta on enemmän reaali maailman semantiikassa ja käsitteellisissä notatioissa. Kuvauksessa esiintyvät suureet ovat esimerkiksi objektit, tapahtumat, elementtien keskinäiset suhteet jne.

Tuonnempana palaamme hieman yksityiskohtaisemmin näihin seikkoihin.

### 5.4 Navigointi ja saanti

Deskriptori schemeihin on sisällytetty myös navigointia ja saantia helpottavia tietoja. Tähän käytetään seuraavia kuvauksia (Beek 2002):

- Yhteenvedoissa kuvaan tiivistetysti median sisältö esimerkiksi haun helpottamiseksi, sisällön visualisoimiseksi tai sonifoimiseksi ("ääntelemiseksi"). Yhteenveto voidaan tehdä rakenteeltaan joko hierarkkiseksi tai sekventiaaliseksi.
- Deskriptorimalleihin voidaan muodostaa näkymiä, jotka pohjautuvat perustuvat sekä partitiointeihin että dekompositiointeihin. Näkymillä voidaan kuvata multimedia-signaalit tilassa, ajassa tai taajuustasolla. Erilaisilla näkymillä sisällöstä on tärkeä osuus esimerkiksi moniresolutio tai progressivisessa haussa ja esittämisessä.
- Varioinneilla muodostetaan tieto eri variaatoiden olemassaolosta kuten erilaisten yhteenvetöjen, skaalojen, resoluutioiden, erikielisten versioiden jne. Esiintymisestä.

### 5.5 Sisällön organisointi

Sisällön organisointi kuvaavien deskriptiomalleilla kuvaillaan sisältökokoelmia. Näitä organisointiin käytettäviä tekijöitä voidaan jaotella esimerkiksi tapahtumien, esiintyvien objektien ja sisällön kuvausten pohjalta (Smith 2002).

## 5.6 Käyttäjän vuorovaikutus

Deskriptori schemeissä käyttäjän vuorovaikutusta kuvaavissa seikoissa käsitellään käyttäjälle asetettavia mahdollisia mieltymyksiä (preferenssejä) ja käyttöhistoriaa. Tällä tavalla voidaan mallintaa esimerkiksi multimedian personoitua saantia, esitystä ja kulutusta (Beek 2002).

## 6 DESKRIPTOREIDEN PERUSELEMENTIT

Luvun 4 MDS kuvauksista tarkastellaan ohessa yksityiskohtaisemmin vain MPEG-7 mediakuvauksen peruselementtejä (muut jätetään lukijan omaehtoisen tutustumisen varaan (hyvänä lähteenä on esimerkiksi Manjunahin kirja (Manjunah 2002)). Peruselementit käsittelevät sellaisia seikkoja kuten sisällön kuvaukset (mukaanlukien ajalliset suureet), median osien väliset linkit ja niiden sijainnit, annotoinnit ja erilaisten sanakirjojen määritykset. Kun muut MPEG-7:n deskriptorit kuvailevat erilaisia vaihtelevia aspekteja multimedian sisällön kuvauksesta niin peruselementit muodostavat yhtenäisen moniin eri sovelluksiin soveltuvan deskriptori-kirjaston (Walker 2002).

### 6.1 Linkitys ja lokalisointi

Linkitys ja lokalisointi menetelmät ovat työkaluja, joilla voidaan viitata MPEG-7 kuvauksiin, liittää kuvaukset mediaan ja kuvaamaan median sisällön ajallisia suureita. Tähän tarkoitukseen joudutaan usein viittamaan deskriptoreihin toisessa tai samassa dokumenteissa. Viitaukset pitäisi voida tehdä joko implisiittisesti (yksikäsitteisen tunnuksen avulla) tai eksplisiittisesti osoittamalla suoraan mediaan (ns. medialokaatorilla).

Usein joudutaan viittamaan yhdestä deskriptorista toiseen ilman että jouduttaisiin rajoittumaan XML:n hierarkiseen dokumenttirakenteeseen tai –rajoihin. MPEG-7 sisältää menetelmät viitata mihin tahansa deskriptorielementtiin kolmella eri tavalla. Ensinnäkin `idref` datatyypillä voidaan viitata saman dokumentin sisällä toiseen deskriptorielementtiin. Toiseksi XML lainatulla `xpath` tyyppillä voidaan viitata deskriptiopuun tiettyyn positioon. Kolmanneksi `href` tyyppillä voidaan viitata suoraan johonkin toiseen deskriptoridokumenttiin (Walker 2002).

Medialokaatorilla voidaan osoittaa suoraan tallennettuun mediaan. Tässä lokaattorin sisältävä deskriptioelementti osoittaa suoraan median sisältöön. Lokaattoreita on standardissa määritelty kolme:

- Geneerinen `MediaLocator`:lla osoitetaan multimediasisältö käyttämällä URI (Uniform Resource Identifier) viittauksia (hieman samantapaan kuin webissä käytetään URL:iä). URI viitteet osoittavat suoraan fyysiseen sijaintiin eivätkä ole median tunnuksia. Mikäli taasen media sisältää useampia mediavirtoja (esimerkiksi erikielisiä audiovirtoja) niin oikean virran valitsemiseksi voidaan suoraan käyttää virtojen inline upotettuja tunnuksia (streamID).
- Temporaaalisella `MediaSegmentLocator`:illa viitataan tiettyyn segmenttiin ajallisesti etenevässä virrassa (esimerkiksi audiossa tai videossa). Tähän voidaan käyttää esimerkiksi ajallisia suureita (aloitushetki ja kesto) tai virran sisältävien tavujen offset (aloitushetki) ja lukumäärä (kesto). Esimerkiksi seuraavanlaista lokatoria voitaisiin käyttää osoittamaan 10 s pitkän pätkän videossa (nimeltä `soccer.mpg`) 3 s päässä aloitushetkestä (Walker 2002):

```
<TemporalSegmentLocator>
```



```

<MediaUri>file:soccer.mpg</MediaUri>
<MediaTimePoint timeBase="../MediaUri">
  PT3S
</MediaTimePoint>
<MediaDuration>PT10S</MediaDuration>
</TemporalSegmentLocator>

```

- ImageLocator kohdistaa yksittäiseen kuvaan tai videoskvenssin frameen. Se on rajoitettu vain tiettyihin ajanhetkiin. Esimerkiksi jos soccer.mpg videossa esiintyy joukkueen logo niin voidaan käyttää ImageLocator:ia inline muodossa ja näin voidaan muodostaa kuvadata suoraan kuvauksen sisään:

```

<ImageLocator>
  <InlineMedia type="image/jpeg">
    <MediaData64>9j/4AAQSkZJRgABAQEAg...</MediaData>
  </InlineMedia>
</ImageLocator>

```

Median paikallistamiseksi hyvin tärkeä suure on myös erilaiset ajalliset määreet. MPEG-7:ssä voidaan määritellä kahdenalaisia ajallisia suureita, ensinnäkin media-aika, joka on median oma aika sekä toiseksi maailmanaika, jolla viitataan yleiseen seinäkelloaikaan. Media-aikoja taasen on määritelty kolme. Yksinkertainen aika, jossa kaikki median tapahtumat esitetään absoluuttisena aikana. Suhteellinen aika, missä tapahtumat esiintyvät suhteessa tiettyyn kanta-aikaan (esimerkiksi uutislähetyksessä voi sisällön sijaintikohta hieman vaihdella, jolloin suhteellinen aika on hyödyllinen suure). Lopuksi inkrementaalisisä ajankuvauksessa lasketaan kesto käyttämällä tiettyjä etukäteen määriteltyjä kiinteämittaisia aikapisteitä (esimerkiksi videossa voi aikapisteiden suuruus olla framien väli 1/25).

## 6.2 Perusmenetelmät

Perusmenetelmät ovat kirjastoituja deskriptiomalleja ja datatyyppejä, primitiivisiä komponentteja, joilla voidaan luoda kompleksisempia ja toimintoihin orientoituneita kuvauksia. Näitä menetelmiä ovat a) erilaiset graafi- ja relaatiomenetelmät, joilla voidaan liittää yhteen kompleksisia kuvausrakenteita, b) erilaiset tekstuaaliset annotaatiot, tekstuaalisten kuvausten muodostamiseksi, c) luokittelumallit ja -termit, joilla voidaan muodostaa ja viitata sanastoihin, d) ihmiset ja paikat, joilla voidaan kuvata multimediasisällössä esiintyvät ihmiset ja paikat, e) affektiiviset kuvaukset, joilla voidaan kuvata tunnevasteita ja lopuksi f) järjestelymenettelyt, joilla voidaan kuvata erilaisia tapoja järjestää deskriptoreita.

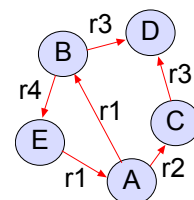
MPEG-7:n deskriptoreiden perusstruktuuri noudattaa yleensä XML-dokumenttien hierarkista sisäkkäistä rakennetta. Kuitenkin monet multimedia sisällön kuvaukset voidaan parhaiten esittää erilaisina graafeina, joissa eri deskriptiomallit on liitetty yhteen erilaisin relaatioin esimerkiksi kun kuva sisältää kerronnallisia tai spatiaalisia elementtien välisiä yhteenkuuluvuuksia:

- Relaatiota kuvaavat deskriptiomallit esittävät yhden tai useamman kuvauksen välisiä suhteita. Relaatiot ovat suunnattuja siten, että siinä on määritelty lähde (source) ja maali (target).
- Graafit taasen esittävät kaavion jossa on lueteltu solmut ja niiden väliset relaatiot. Jokaista graafin solmua kohden esiintyy omalla attribuutillaan identifioitava Node elementti. Oheiseen esimerkkiin on hahmoteltu yksinkertainen graafi:

```

<Graph>
  <Node id ="A" /> <Node id ="B" />
  <Node id ="C" /> <Node id ="D" />
  <Node id ="E" />

```



```

<Relation type "#r1" source="#A" target="#B">
<Relation type "#r2" source="#A" target="#C">
<Relation type "#r3" source="#B" target="#D">
<Relation type "#r4" source="#C" target="#D">
<Relation type "#r5" source="#B" target="#E">
<Relation type "#r6" source="#E" target="#A">
</Graph>

```

Multimedian sisällön kuvaus käyttäen luonnollista kieltä kutsutaan tekstin annotoinniksi. Käyttämällä tekstin annotointia hakuun, luettelointiin ja indeksointiin on ollut perinteiltään pitkä ja sovelluksia on ollut käytössä laajasti. MPEG-7 tukee tätä perinnettä TextAnnotation datatyypillä. Datatyyppi sallii vapaan tekstin (FreeTextAnnotation), avainsana (Keyword), strukturoidun (StructuredAnnotation) ja riippuvuuksien mukaisen (DependencyStructure) annotoinnin:

- Vapaassa tekstin annotoinnissa sisältö kuvataan luonnollisen kielen lauseilla. Esimerkiksi ”Futiksessa Suomi tekee maalin Ruotsiin. Maalin teki Litmanen”:

```

<TextAnnotation>
  <FreeTextAnnotation xml:lang="fi">
    Futiksessa Suomi tekee maalin Ruotsiin.
    Maalin teki Litmanen.
  </FreeTextAnnotation>
</TextAnnotation>

```

- Vapaa tekstin annotointi on toki inhimilliselle lukijalle helppo mutta koneellisessa luennassa ei aina ole näin. Kone voi ymmärtää paremmin avainsanoja, vaikka alkuperäisen lauseen rakenne näin häviää:

```

<TextAnnotation>
  <KeywordAnnotation>
    <Keyword>Futis</Keyword>
    <Keyword>Suomi</Keyword>
    <Keyword>Maali</Keyword>
    <Keyword>Ruotsi</Keyword>
    <Keyword>Litmanen</Keyword>
  </KeywordAnnotation>
</TextAnnotation>

```

- Rakenteellisessa annotoinnissa taas pyritään muodostamaan tiettyjä rakenteita. Tässä normaalisti pyritään muodostamaan kysymyksiin kuka, mikä objekti, mikä tapahtuma, missä, koska miten ja kuinka vastauksia. Esimerkiksi:

```

<TextAnnotation>
  <StructuredAnnotation>
    <Who><Name>Suomi</Name></Who>
    <WhatAction><Name>Futis maali</Name></WhatAction>
    <Where><Name>Lyon, Ranska</Name></Where>
    <When><Name>Maaliskuu 25, 1998</Name></When>
  </StructuredAnnotation>
</TextAnnotation>

```

- Riippuvuusrakenteinen annotointi on ehkä mutkikkain annotointi tapa. Se on erittäin tehokas jos haluamme esittää annotoinnin lingvistisen rakenteen. Menettely perustuu ns. riippuvuuksia kuvaavaan kielioppiin, jolla kuvataan lauseen syntaktinen rakenne sanojen välisinä riippuvuuksina. Rakenne kuvataan eräänlaisena puumaisena rakenteena.

Deskriptiomallit muodostavat siis keskeisen tekijän multimedian sisällön kuvaamisessa. Kuitenkin vähintään yhtä tärkeä deskriptoreissa käytettävä sanasto. Lisäksi, jotta eri järjestelmät voisivat ymmärtää toisiaan pitää tämä sanasto myös standardoida. Eri järjestelmissä pitäisi voida käyttää omia sanastojaan ja laajentamaan olemassa olevia sanastoja. Koska sovelluksia on erittäin paljon niin on melko selvää, että MPEG-7:ssä ei voida kaikkia sanastoja standardoida. Ongelman ratkaisemiseksi MPEG-7:ssä on määritelty

tyyppi `ClassificationScheme`, jonka avulla voidaan määrittellä erilaisia sanastoja ja termien muodostamia alueita. `ClassificationScheme` deskriptiomalli sisältää listan termien määrittelyistä ja mahdollisesti mukaanotettujen toisten `ClassificationScheme`-jen. Jokainen `ClassificationScheme` on yksikäsitteisesti identifioitu URI:lla. Valinnaisella `domain` attribuutilla voidaan kuvata se joukko deksriptoreita, joihin tiertty datatyyppi voidaan soveltaa (Walker 2002). Oheinen esimerkki kuvaa miten `ClassificationScheme` voidaan käyttää urheilutermin määrittelyyn, tämä on nimetty URI attribuutilla `urn:examples:cs:SportsCS`, termit on nimetty suomeksi ja englanniksi mutta termin tunnus (`termID`) on kieliriippumaton:

```
<ClassificationScheme uri="urn:examples:cs:SportsCS">
  <Term termID="soccer">
    <Name xml:lang="en">Soccer</Name>
    <Name xml:lang="fi">Futis</Name>
  </Term>

  <Term termID="Basketball">
    <Name xml:lang="en">Basketball</Name>
    <Name xml:lang="fi">Koripallo</Name>
  </Term>

  <Term termID="baseball">
    <Name xml:lang="en">Baseball</Name>
    <Name xml:lang="fi">Pesäpallo</Name>
  </Term>

  <!--muut urheilutermit -->
</ClassificationScheme>
```

MPEG-7:ssä ihmisiä (esimerkiksi näyttelijöitä), organisaatiota (esimerkiksi yritykset) ja ihmisten muodostamia ryhmiä (esimerkiksi orkesteri) kutsutaan nimellä *agentti*. Agentin sanotaan olevan suure, joka on vastuussa toiminnasta. `Person` deskriptorimalli kuvaa joko todellista tai kuvittellista henkilöä, jolla on nimi, liittyy organisaatioon tai ryhmään ja jolla on osoite. `PersonGroup` on taasen ihmisryhmää kuvaava deskriptorimalli, jolla on kollektiivinen nimi ja jonka jäsenet ovat helposti tunnistettavissa (esimerkiksi musiikkiryhtye). Nimen lisäksi voidaan ryhmän tyyppi määrittellä (esim. rock yhtye), yhtyeen jäsenet jne. Lopuksi `Organization` deskriptorimalli kuvaa organisaatiota eli joukkoa ihmisiä, joilla on kollektiivinen nimi ja jonka jäsenet eivät ole kovin helposti tunnistettavissa tai joiden jäsenmäärä on erittäin suuri. Oheinen esimerkki kuvaa suomen jalkapallomaajoukkuetta ja sen jäseniä:

```
<PersonGroup>
  <Name>Suomen jalkapallo maajoukkue </Name>
  <Kind><Name>Futisjoukkue</Name></Kind>
  <Member xsi:type="PersonType">
    <GivenName>Jari</GivenName>
    <FamilyName>Litmanen</FamilyName>
  </Member>
  <!--lisää jäseniä tähän -->
</PersonGroup>
```

Paikkoja kuvaava `Place` deskriptorimalli kuvaa olemassa olevia historiallisia tai kuvitteellisia paikkoja. Niitä käytetään kuvaamassa paikkoja, joita on esitetty median sisällössä (esimerkiksi historialliset tai kuvitteelliset lavasteet ja paikat) ja paikkoja todellisessa maailmassa (esimerkiksi kuvauspaikka). Malliin sisältyy nimi, rooli ja maantieteellinen sijainti. Oheinen esimerkki esittää espanjalaista jalkapallostadionia Riazoria:

```
<Place>
  <Name xml:lang="es">Estadio de Riazor </Name>
  <Name xml:lang="fi">Riazor Stadion</Name>
  <Role><Name>Futis peluupaikka</Name></Role>
  <Region>es</Region>
</Place>
```

Tunteita ja mielialoja voidaan kuvata perustyökaluihin kuuluvalla *Affective* deskriptiomallilla. Tässä asetetaan tietty lukuarvo jokaiseen deskription osaan (esimerkiksi videosekvenssiin. Jokainen lukuarvo ilmaisee tietyn tunteen (kiihottuneisuus, onnellisuus, viha jne.) suhteellisen intensiteetin (Walker 2002). Esimerkiksi jos olemme mitanneet tietyn futismatsin katsojan kiihottuneisuudenasteen pelin tai videon edistyessä voimme muodostaa seuraavanlaisen kuvauksen:

```
<Affective>
  <Type><Name>Excitement</Name></Type>
  <Score idref = "Seg0">0.00360117</Score>
  <Score idref = "Seg1">0.825006</Score>
  <Score idref = "Seg2">-0.20360117</Score>
  ;
</Affective>
```

## 7 VISUAALISET DESKRIPTORIT

MPEG-7:n visuaalisten deskriptoreiden tavoitteena on muodostaa standardoitu kuvaus videovirroissa tai still-kuvissa esiintyvistä visuaalisista tai graafisista suureista. Kuvauksen avulla helpotetaan käyttäjiä tai sovelluksia tunnistamaan, luokittelemaan tai suodattamaan kuvia ja videoita (Manjunath 2002). Näillä alatasen deskriptoreilla voidaan vertailla, suodattaa tai selata kuvia ja videota puhtaasti ei-tekstuaalisella sisällön visuaalisilla kuvauksilla usein yhdessä tekstuaalisten kuvausten kanssa.

Sovellusalueita on monia esimerkiksi lääketieteellinen kuvien arkistointi, kuvakirjastot, TV-kanavien valinta ja multimedia editointi. Eri sovelluksissa voitaisiin tehdä seuraavanlaisia toimenpiteitä (Manjunath):

- Grafiikka: Piirretään muutama viiva kuvaruudulle ja vasteena saataisiin vastaavanlainen kuva tai logo.
- Still kuva: Kerrotaan haluttu objekti (mukaan lukien väriläikkä tai tekstuuri) ja vasteena saadaan esimerkkejä kiinnostavista kohteista.
- Video: Kuvataan annettujen video-objektien liikkeitä, kameran asentoja tai objektien välistä suhteita saadaan vasteena luettelo videoista joilla on samanlaiset tai eriävät temporaaliset suhteet.
- Video aktiviteetti: Kuvataan videossa esiintyvä aktiviteetti (esimerkiksi jalkapallossa maalin syntytilanne) ja saadaan vasteena luettelo videoista joissa tapahtuu samanlainen aktiviteetti.

Visuaaliset deskriptorit kuvailevat median sisältöä perustuen visuaaliseen informaatioon, esimerkiksi objektien muotoon, kokoon, liikkeisiin tai kameran asentoihin. MPEG-7:n visuaaliset deskriptorit voidaan karkeasti luokitella yleisiin visuaalisiin deskriptoreihin ja alue spesifeihin deskriptoreihin. Ensin mainittuihin sisältyy väri, tekstuuri, muoto ja liike deskriptorit. Jälkimmäisiin taasen kuuluvat sovelluskohtaiset deskriptorit mukaan lukien kasvojen tunnistus.

Väri on yksi yleisimmin käytetyistä visuaalinen piirre kuvien ja videon haussa ja kuvauksissa. Väripiirteet ovat suhteellisen robusteja katselukulman vaihtelun, translationaalisen ja rotationaalisen siirtymän suhteen. Standardi sisältää kuusi väri deskriptorita, jotka kuvaavat eri piirteitä mukaan lukien värijakauma, värin spatiaalinen layout ja rakenne (Ohm 2002).

Tektuurilla tarkoitetaan visuaalista hahmoa, joka voi olla homogeeninen ja on tulos monien värien ja näiden intensiteettien esiintymisestä kuvassa. Lähes kaikkien pintojen ominaisuus on niiden tekstuuri esimerkiksi pilvet, puu, kaarna ja kangas. Tekstuuri sisältää tärkeää rakenteellista tietoa ja niiden suhteet ympäristöönsä. Kuvaamalla tekstuureja MPEG-7 tekstuuri-deskriptoreilla voidaan muodostaa tehokkaita hakuja sekä homogeeniselle että epähomogeeniselle tekstuureille (Choi 2002).

Monissa kuvakanta-sovelluksissa muodostavat kuvan esittämät muodot tärkeän vihjeen samankaltaisuuden löytämiseksi. MPEG-7:ssä on sekä alue että ääriviiva-deskriptorit määritelty. Sovelluksina on esimerkki liikemerkkien ja binääristen kuvien kuten myös 2-D ja 3-D objektien haku (Bober 2002).

Kuvaamalla liikettä voidaan em. visuaalisten deskriptoreiden lisäksi parantaa olennaisesti hakuja videovirroista. MPEG-7 sisältää lukuisan joukon deskriptoreita, joilla voidaan kuvata objektien ja kameran liikettä yhteneväisellä ja tehokkaalla tavalla (Jeanin 2002).

## 7.1 Kasvojen tunnistus

MPEG-7:ään sisältyy deskriptorit kasvojen tunnistamiseksi. Kasvojen tunnistamisen sovellusalueet ovat laajat esimerkiksi kuvan ja videon haku tietokannoista ja turvallisuussovellukset. Standardointiin ehdotettiin useita eri menetelmiä, joista valittiin ns. Principal Component Analysis (PCA). Tässä kasvojen piirteistä muodostetaan 48 piirrevektoria, joita käytetään deskriptoreiden perusteena. Piirrevektorit on johdettu luokittimien opetuksessa käytettyjen kasvojen piirteiden ominaisvektoreista. PCA on kohtalaisen robusti katselukulma ja valaistusolosuhde muutoksille.

## 7.2 Väri

Väri on keskeinen visuaalinen attribuutti kaikessa inhimillisessä näkemisessä ja kuvien tietokonetuetussa prosessoinnissa. Värejä varten MPEG-7:ssä on määritelty 7 deskriptoriryhmää:

- Väriavaruus (Color Space) deskriptorilla voidaan valita kuvauksessa käytettävä värimalli. Väriavaruuksina tunnetaan MPEG-7:ssä seuraavat RGB, luminanssi-krominanssi-signalit YCbCr (ITU-R 601 mukaisesti), HSV (hue, saturation, value) ja värien havaitsemiseen läheisesti liittyvä HMMD (hue, max, min, diff) -avaruus. Diskreetisti jatkuvan väriavaruuden valinnan lisäksi standardi mahdollistaa erilaiset värien kvantisoinnit.
- Määrävä väri (Dominant Color) deskriptorilla on mahdollista määritellä pieni määrä (kuvaa, videota) dominoivia värejä ja näille sellaisia tilastollisia ominaisuuksia kuten jakauma ja varianssi. Deskriptorilla pyritään esittämään tietyn kuvan sisältyvän alueen värit tehokkaasti, kompaktisti ja intuitiivisesti.
- Skaalautuva väri - deskriptori (Scalable Color) muodostetaan sellaisen kuvan väri histogrammeista, jotka on laskettu kiintesästi kvantisoidusta HSV-esityksen mukaisesta kuvasta.
- Ryhmä-deskriptori (Group of Frame tai Group of Pictures) taas on laajennus em. skaalattavasta väri-deskriptorista. Tällä voidaan kuvata kokonaisten ryhmien tai videosekvenssien väriominaisuuksia.
- Värirakenne (Color Structure) deskriptori perustuu myös värihistogrammeihin, mutta se kohdistuu pieniin lokalisoituihin kuva-alueisiin. Deskriptoria voidaan käyttää vain HMMD-väriavaruudessa.

- Värien layout deskriptorilla kuvataan värien spatiaalisia ulottuvuuksia. Tämä tehdään jakamalla tietty kuvan osa hilamuotoisiin osa-alueisiin, joille kullekin tehdään diskreetti kosinimuunnos. Muunnoskertoimet esitetään sen jälkeen deskriptorissa.

### 7.3 Tekstuuri

Värien lisäksi toinen tärkeä kuvien ja videon luonnehtiva ominaisuus on tekstuuri, pinnan rakenne. Kuvien selaamisessa ja haussa tekstuuri on tärkeä primitiivi samankaltaisuutta luonnehtivissa hahmoissa. Esimerkiksi kuvat, jotka esittävät vettä, kangasta, kukkia jne. voidaan intuitiivisesti erottaa toisistaan joten niiden pistäisi myös kuulua MPEG-7 kuvauksiin. Tyypillisesti tekstuurit ovat alueellisia, rajallisia, ominaisuuksia. Monen kuvan voidaan ajatella muodostuvan jonkinlaisesta mosaiikista homogeenisia tekstuureja. Tällaista tekstuurirakennetta voidaan käyttää esimerkiksi kuvien indeksointiin. Esimerkiksi selailtaessa ilmakuvakantaa voi joku olla kiinnostunut pysäköintialueista eri kuvissa. Tyypillisesti tällaiset alueet ovat muodostuneet autoista, jotka on pysäköity säännöllisiin intervaleihin. Kaukaa katsottaessa tällainen alue muodostaa selkeän tekstuurin.

Toistaiseksi MPEG-7 sisältää 3 erilaista tekstuurin deskriptoria:

- Homogeenisella tekstuurin deskriptorilla muodostetaan 62 luvulla kvantitatiivinen esitys tekstuurista. Luvut kuvaavat tekstuurin energiaa ja tämän jakaumaa joukossa (spatiaalisia) taajuuskanavia. Nimensä mukaisesti deskriptorilla voidaan luonnehtia homogeenisia alueita ja on myös tehokas samankaltaisuuksien havaitsemisessa.
- Tekstuurin selaus deskriptori (texture browsing descriptor) soveltuu hyvin erilaisten tekstuurien silmä määräiseen selaamiseen. Deskriptori on sängen kompakti, se vaatii korkeintaan 12 bittiä. Näillä voidaan tekstuurista esittää samankaltaisia piirteitä kuin mitä ihmissilmä havaitsee eli tekstuurin rakeisuutta, säännöllisyyttä ja suuntaominaisuuksia.
- Reuna histogrammi deskriptorilla kuvataan kuvaan sisältyvien reuna-alueiden spatiaalista jakaumaa. Se on tehokas mitta sovittamaan osittain vaihtuvaa epäyhtenäistä tekstuuria. Erityisen hyvin se soveltuu kuviin jotka ovat epähomogeenisesti teksturoitu, kuten luonnoksiin ja luontokuviin (eläimet ja kasvit).

### 7.4 Muoto

Kohteen muoto on ehkä yksi merkittävimmistä vihjeistä minkä mukaan ihmiset erottavat esineitä toisistaan. Osittain tämä johtuu siitä että kappaleiden toiminnallisuus usein liittyy niiden muotoon (vrt. esimerkiksi autoa ja polkupyörää keskenään). Monissa sovelluksissa muoto on keskeinen myös valintaparametri. Esimerkiksi e-commerce sovelluksissa pelkällä tekstuaalisella kuvauksella on hyvin vaikea tehdä hakuja ja ostopäätöksiä. Query-by-example (QBE) sovelluksissa tehdään hakuja näyttämällä esimerkkejä, antamalla visuaalisia vihjeitä. QBE:n luotettavuus voisi olla hieman kyseenalaista ilman selkeästi määriteltyjä muotomääreitä.

MPEG-7:ään on sisällytetty 3 erilaista deskriptoria muodon määrittämiseksi, näistä 2 on suuntautunut 2-D kuviin ja yksi 3-d polygonimalleihin:

- Alue perustaiset muodon deskriptorit kuvaavat pikseleiden jakaumista 2-D objektin puitteissa. Koska deskriptori perustuu sekä reunoihin sekä näiden rajoittamiin sisäisiin pikseleihin soveltuu se myös ei-kytkettyihin objekteihin tai muotoihin, jotka sisältävät reikiä. Deskriptorit perustuvat ns. Angular Radial Transform muunnokseen.

- Suljettuihin alueisiin (contour) pohjautuvat deskriptorit ovat erittäin kompakteja kuvauksia ja ovat myös robusteja muodon muunnosten, deformaatioiden, suhteen. Deskriptorit emuloivat myös sangen hyvin ihmisen havaintojärjestelmää [BOBE02].
- 3D muoto deskriptoreilla esittävät luonteenomaisia piirteitä objekteista, joita voidaan kuvata 3-D polygonirakenteilla ja -verkoilla. Deskriptorit perustuvat lokaaleista geometrisista piirteistä laadittuihin histogrammeihin. Näin kuvaukset ovat robusteja erilaisten affiinien muunnosten suhteen.

## 7.5 Liike

Liikepiirteet videosekvenssissä muodostavat ehkä keskeisimmät temporaaliset dimensiot ja ovat tämän vuoksi keskeisiä tekijöitä videon luokittelemisessa. Liikkeen kuvaamiseen on MPEG-7:ssä varattu 4 deskriptoria. Nimittäin deskriptorit, jotka kuvaavat yleisiä liikeaktiiviteetteja, kameran liikettä, liikeratoja sekä parametrisoituja liikkeitä.

## 8 JOHTOPÄÄTÖKSET

MPEG-7 on osoittautunut vitsiksi multimedia datan kuvaamisessa. Sitä voidaan soveltaa lähes kaikkialla siellä missä AV dataa voidaan esittää digitaalisesti. Sen ominta sovellusalueita ovat erilaiset tietokannat, joissa haut voidaan tehdä sisällön mukaisin kuvauksin ja streamasussovellukset, joissa dataan voidaan lisätä erilaisia luonnehtivia lisämääreitä kuten metadata (tekijät, nimikkeet, luontipäivät jne.), käyttöoikeudet ja kompositioinnit peruselementeistä tai muista läheisesti liittyvistä seikoista.

Toisaalta monet MPEG-7:ään sisältyvät kuvaukset ovat melko kompleksisia ja edellyttävät varsinaisen tekstuaalisti (interaktiivisesti) laadittujen kuvausten lisäksi erilaisia työkaluja mediaan sisältyvien hahmojen, äänten ja muiden piirteiden automaattista tai puoliautomaattista irrottamista. Standardi ei käsittele miten tällainen irrotus kannattaisi tehdä eikä se myös käsittele siten miten kuluttajan päässä irrotettuja piirteitä pitää hyödyntää. Nämä ovat sovelluksen asioita.

MPEG-7 muodostaa myös kiinteän ja yhteisesti hyväksytyyn alustan, jonka pohjalta voidaan laatia erilaisia sovellusspesifejä suosituksia ja ohjeita. Tämän syyn vuoksi kaikissa uusissa sovelluksissa, joissa käsitellään multimediaa oli se sitten still kuvia, videota, ääntä tms. olisi syytä ottaa huomioon MPEG-7 kuvaustavat. MPEG-7 on siis tullut jäädäkseen.

Kirjallisuutta:

Suurin osa lähteinä käytetyistä materiaalista on löydettävissä artikkelikokoelmasta:

- [MANJ02] Manjunah B.S, Salember P., Siokora T (ed): Introduction to MPEG-7, multimedia content description interface. John Wiley & Sons. Ltd, Englans 2002
- [AVA02] Avaro O., Salember P.: System Architecture, in [MANJ02]  
[BEEK02] van Beek P., Smith J.: Navigation and Summarization, in [MANJ02]  
[BEEK02a] van Beek P., Yoon K., Ferman A.: User Interaction, in [MANJ02]  
[BENI02b] Benitez A., Martinez J., Rising H., Salember P.: Description of a Single Multimedia Document, in [MANJ02]
- [BOBE02] Bober M., Preteus F., Kim W-Y.: Shape Descriptors, in [MANJ02]  
[CAS02] Casey M.: Sound Classification and Similarity, in [MANJ02]  
[CHOI02] Choi Y., Won C., Ro Y., Manjunath B.: Texture Descriptors, in [MANJ02]  
[CHAR02] Charlesworth J., Garner P-: Spoken Content, in [MANJ02]  
[DAY02a] Day N., Search and Browsing, in [MANJ02]  
[DAY02b] Day N., Sekiguchi S., Sasaki M.: Mobile Applications, in [MANJ02]  
[HUNT] Hunter J., Armstrong L.: A Comparison of Schemas for Video Metadata Representation  
[HEUE02] Heuer J., Thienot C., Wollborn M.: Binary Format, in [MANJ02]  
[HUNT02] Hunter J., Seuýrat C.: Description Definition Language, in [MANJ02]  
[JEAN02] Jeannin S., Divakaran A., Mory B.: Motion Descriptors, in [MANJ02]  
[LIND02] Lindsay A., Burnett I., Quackenbush S., Jackson M.: Fundamentals of Audio Descriptors, in [MANJ02]
- [LYMA02] Lyman P., Varian H.R.: How Much Information, 2000, <http://www.sims.berkeley.edu/how-much-info>
- [MANJ02] Manjunath B., Sikora T.: Overview of Visual Descriptors, in [MANJ02]  
[OHM02] Ohm J-R., Ciepinski L., Kim H., Krishnamachari S., Manjunath B, Messing D., Yamada A.: Color Descriptors, , in [MANJ02]
- [SALE02] Salember P., Smith J.: Overview of Multimedia Description Schemes and Schema Tools, , in [MANJ02]
- [SMIT02] Smith J., Benitez A.: Content Organization, , in [MANJ02]  
[WALK02] Walker T., Heuer J., Martinez J: Basic Elements, , in [MANJ02]