

Janne Lilleberg ja Kaarle Kannisto

AV-alan metadatan yhteiskäyttö

Selvitys 2021

AUDIOVISUAL
PRODUCERS
FINLAND
APFI

Sisältö

Tiivistelmä	4
AV-alan toimijat	5
Katsojat	5
Mainostajat	5
AV-Tuottajat	6
Tekijät	6
Tekijänoikeuden yhteishallinto-organisaatiot	7
Kanavat	7
Striimaajat	7
Teatterit ja festivaalit	8
Riskirahoittajat	8
Agentit/levittäjät	9
Tuottajakuluttajat (prosumers) ja nettisidosryhmät	9
Täydentävät palvelut	9
Suomen elokuvasäätiö SES	9
Finnpanel	10
Lehdet ja verkkopalvelut	10
Kansallinen audiovisuaalinen instituutti KAVI	10
AV-ekosysteemin data	11
Datan synty	11
Datan tyypittely	12
Datan arvo ja merkitys	13
Tekijänoikeusjärjestelmän kannalta keskeinen data	14
Teosdata	14
Data oikeuksista	15
Data AV-teoksen käytöstä	18
Lineaari-TV	18

Striimauspalvelut	19
Konteksti-sidonnainen data	20
Tuottaja-portaali -ratkaisu akuuttiin tarpeeseen	21
Keskitetty datan syöttö ja jakelu	21
Tunniste	22
Tuottaja-portaalin demo	23
Tarvittavan datan määrittely	24
Toimintatapojen muutos ja sen vaatima kouluttaminen	24
Ratkaisun kustannukset	24
Turvallisuus	25
Hallinto	25
Toimialan näkymät	26
Tekijänoikeuksien luovutusten standardoimattomuus	26
Lohkoketjut	26
Data-intensiteetin kasvu ja avoin data	27
Valinta/suosittelualgoritmien kehittäminen	27
Lainsäädännön vaikutukset	28
Kansainväliset toimijat	29
Tuottajakuluttajat	29
Sisältöjen laitton käyttö, piratismi ja mikromaksaminen	29
Legacy tuotteet	30
Yksittäisen ohjelman osat	30
Projektin jatkoaskelmat ja rajoitteet	31
Yhteenveto	31
Haastatellut henkilöt	33
Liite 1: Portaalin tekninen kuvaus	34

Tiivistelmä

Selvitimme OKM:n rahoituksella tekijänoikeusjärjestelmän kannalta keskeisen metadatan syntyminen, tallentamisen, siirtymisen eri organisaatioiden välillä ja tuon datan hyödyntämisen AV-alalla. Määrittelimme AV-alan / ekosysteemin kattavan AV- sisällön tuottamisen ja jakelun, sisällön tuottamiseen ja jakeluun liittyvästä tekniikasta ja toimialaa säätelevästä regulaatiosta. Tutkimus tehtiin haastattelemalla 24 alan ammattilaista keväällä 2021.

Suomalainen AV-ala on muutoksessa. Muutosta aiheuttaa erityisesti kansainväliset sisältöpalvelut, kuten Netflix ja HBO, ja sisällönvälitysalustat, kuten Youtube ja Facebook. Kilpailu mainostajien euroista ja katsojista kiristyy. Toisaalta Suomessa tuotettu sisältöä myydään kansainvälisesti enemmän kuin koskaan ennen.

Sisältöpalvelut ja sisällönvälityspalvelut keräävät määrätietoisesti tietoa käyttäjistään ja heidän toimistaan. Tätä tietoa voidaan käyttää paitsi mainonnan kohdistamisessa, myös sisällön ostamisessa ja kehittämisessä. Tämä auttaa kilpailtaessa katsojien ajasta.

Suomalainen AV-ala jakaa tietoja tällä hetkellä arastellen. Tällä hetkellä data on suurelta osin suljetuissa tietokannoissa tekijänoikeusjärjestöissä sekä lähettäjä- ja striimausyrityksissä. Avointa dataa on hyvin vähän. Yhteisiä tunnisteita ja luokitteluita ei ole laajasti käytössä.

Jotta dataa teoksista voisi jakaa tehokkaasti, ensimmäinen edellytys on tietysti tahto jakaa tietoa. Jokaiselle mukanaolijalle täytyisi olla hyötyä olla mukana. Jos suomalaiset toimijat ajattelevat kilpailijoinaan ennen kaikkea toiset suomalaiset toimijat, saattaa näyttää siltä, että dataa ei kannata jakaa. Jos kriittisinä kilpailijoina pitääkin kansainvälisiä toimijoita, näkökulma muuttuu.

Jakamisen täytyy tuottaa tehokkaasti laadukasta dataa, joka edellyttää teokselle uniikkia ID. Näin eri lähteistä tulevaa dataa voi yhdistää. Globaalisti uniikki, maksullinen ID ei ole tarkoituksenmukainen kaikelle sisällölle. Lähetysyhtiöiden productID voisi toimia jakamisen mahdollistavana ID:nä.

Pitemmällä aikavälillä datan monenkeskiselle jakamiselle täytyy sopia pelisäännöt, jossa auttaa esimerkiksi Sitran reilun datatalouden sääntökirja. Datatalous tarkoittaa toisaalta olemassa olevissa yrityksissä datan nostamista keskiöön, jotta kilpailukykyä voidaan parantaa. Toisaalta se tarkoittaa aivan uusien yritysten syntymistä, jotka yhdistävät ja analysoivat dataa niin, että he luovat lisäarvoa muille toimijoille.

Ensimmäisenä askelmana voitaisiin ottaa käyttöön tuottajille portaali datan välittämiseksi yhteishallinto-organisaatioille ja kanaville. Mallina voitaisiin käyttää mainosten välittämisen portaalia spotgatea, jota hallinnoi kaupallisten TV-kanavien Screenforce.

Kuvaamme ensin AV-alan toimijat korostaen erityisesti heidän vaikutustansa toimialan datan ja tiedon syntyamiseen ja tarpeeseen. Sen jälkeen kuvaamme AV-alan dataa, keskittyen tekijänoikeusjärjestöjen kannalta keskeiseen dataan. Kuvaamme havaitsemamme lyhyen aikavälin tarpeet ja ratkaisuksi ehdottamamme yhteisen tiedonkeruun portaalin. Lopuksi tiivistämme havaintojamme haastatteluissa esiin tulleista yksityiskohdista, jotka on hyvä pitää mielessä jatkotyön kannalta.

AV-alan toimijat

Esittelemme seuraavassa lyhyesti datan kannalta keskeisimmät AV-alan toimijat. Jätämme tekniikan vaikutukset huomiotta, ellei tekninen ratkaisu vaikuta keskeisesti liiketoiminnan logiikkaan tai datan syntymiseen ja siirtymiseen.

Katsojat

Katsojat, yleisö, kuluttajat rahoittavat AV-sisältöjen tuotannon ja jakelun joko suoraan maksamalla sisällöstä tai välillisesti mainonnan tai verovarojen kautta. Suoraan sisällöstä maksetaan esimerkiksi maksamalla kanavan kuukausimaksua, ostamalla pääsylipun elokuviin tai ostamalla tai vuokraamalla tallenne. Mainostajilla tavoitteena on saada aikaan myyntiä ja sen vuoksi halutaan saavuttaa oikea asiakasryhmä eli kohdeyleisö tarkoituksenmukaisella markkinointiviestillä. Valtiolliset kanavat toteuttavat julkista palvelutehtäväänsä.

Liikkuvaa kuvaa käytetään enemmän kun koskaan. Lineaarisen TV:n katsominen on pysynyt suunnilleen samana. Verkkosisällöt tarjoavat uusia ominaisuuksia katsojille, joka on johtanut kulutuksen kasvuun ja kokonaiskatsomisen lisääntymiseen. Television katsominen on edullinen ja helppo vapaa-ajanviettotapa.

Kuluttaja haluaa löytää helposti hänelle juuri nyt sopivaa katseltavaa. Tähän hänelle tarjotaan erilaisia ratkaisuja: Lineaari-TV lupaa tarjota sopivaa sisältöä valitsemalleen kohdeyleisölleen. Verkkotallennuspalvelut ja oma tallennuslaite mahdollistavat valittujen teosten katsomisen oman aikataulun mukaan. Striimauspalvelut suosittelvat aiemman katsomisen ja muiden katsojien käyttäytymisen perusteella uusia sisältöjä katsojalle. Striimaus tarkoittaa tässä selvityksessä kaikkea katsomisajasta riippumatonta sisällön jakamista eli tilaukseen perustuvat palvelut kuten Netflix, ilmaiset palvelut kuten Areena ja ilmaiset video-on-demand-palvelut.

Keskimääräinen amerikkalainen katsoja käyttää vuodessa 45 tuntia miettiessään mitä katsoisi seuraavaksi. Kuluttaja käyttää valintansa tueksi mm. tuttujensa suosituksia, omaa etsivätyötänsä verkossa, palveluntarjoajien suosituksia ja mainoksia. Valinta ei ole helpottumassa, koska tapoja ja lähteitä katsoa sisältöjä tulee koko ajan lisää.

Katsoja arvostaisi paremmin osuvia suosituksia, helpompaa kiinnostavien sisältöjen löytämistä ja helpompaa mahdollisuutta katsoa noita sisältöjä. Laadukkaan datan saatavuus auttaa näissä.

Mainostajat

Suurin osa ilmaiseksi vastaanotettavista televisiokanavista Suomessa on rahoitettu mainoksilla. Mainostajat yrittävät sijoittaa mainosrahansa mahdollisimman tehokkaasti omaa viestiänsä edistävästi. Mainostajalle on keskeisintä tavoittaa oikea kohdeyleisö oikeanlaisella viestillä.

Mainostaja siis tarvitsee tietoa katsojapotentiaalista ennen kaikkea kohdeyleisössään. Lisäksi he haluavat jälkikäteen verifikaation tavoitavuudesta ja kontaktien laadusta.

Mainoksien logistiikka Suomessa toimii kaupallisten TV-kanavien Screenforce Ry:n hallinnoiman Spotgate-järjestelmän kautta. Siinä jokaiselle mainokselle annetaan ID.

AV-alan lisääntynyt data tuo etuja mainostajille useaa kautta. Kanavilla on enemmän työkaluja ohjelmistosuunnitteluun ja näin saavutetaan uusia ja/tai tarkemmin määriteltyjä kohdeyleisöjä. Lisääntynyt data katsojista ja heidän käyttäytymisestään mahdollistaa kohderyhmien automaattisen tunnistamisen ja tehokkaamman mainonnan kohdentamisen.

AV-Tuottajat

Tuottajat tuottavat AV-sisällön. He siis etsivät ja kehittävät kiinnostavia käsikirjoituksia / formaatteja, hankkivat tarvittavat immateriaalioikeudet, rahoituksen, palkkaavat henkilöt ja hankkivat muut resurssit, varmistavat toteutuksen sekä valvovat tekijänoikeuksien toteutumista. Tuottajalla on taloudellinen vastuu projektista. Suomessa TV-sarjoissa rahoitus usein hankitaan myymällä ohjelmaidea etukäteen kanavalle. Draaman ja elokuvan tekeminen on kalliimpaa ja tuotantoprojekti kestää ajallisesti pidempään, joten siihen joudutaan hakemaan monipuolisempaa rahoituspakettia.

Tuottaja on myös pääsääntöisesti teosten oikeuksien hallinnoija. Tuottaja varmistaa, että sillä on tarvittavat sopimukset tekemiseen osallistuneiden henkilöiden kanssa ja muun hyödynnetyn immateriaalioikeuksien käyttöön (musiikki, kirja, formaatti, hahmo). Jos teos on myyty määräajaksi ja/tai määrätylle maantieteelliselle alueelle, tuottaja hallitsee teoksen hyödyntämistä määräajan jälkeen ja/tai muilla maantieteellisillä alueilla joko itse tai käyttämällä agenttia/levittäjää. Tuottaja tai kanava myös pyrkii estämään laittoman käytön.

Tuottaja on keskeisin tiedon synnyttäjä. Tämän tiedon välittäminen eri osapuolten välillä AV-ekosysteemissä on edelleen laajasti manuaalista. Tietoa välitetään sähköpostilla tai syöttämällä tietoa verkkolomakkeisiin. Määrämuotoista tietojen siirtoa on vähän. Vakiintuneita tunnisteita ei ole.

Tuottaja saa monenlaisia hyötyjä AV-datan laajemmasta jakamisesta. Ohjelmalogistiikka toimijoiden välillä helpottuu ja siirrettävän tiedon laatu paranee. Tieto tekijänoikeuksista saadaan tarpeellisilta osin ko- neluettavaan muotoon. Automaatiota eri alueilla voidaan lisätä. Datan yhdistely avaa uusia analyysimahdollisuuksia. Validoidun datan läpinäkyvyys saattaa avata uusia rahoitustapoja tuotannoille.

Tekijät

Kaikki tekijänoikeudet syntyvät luovan panoksensa antaneille henkilöille. Tässä selvityksessä nimitämme heitä kaikkia tekijöiksi, emme siis puhu erikseen tekijöistä ja esiintyjistä. Koko AV-toimiala perustuu näiden oikeuksien hyödyntämiseen.

Tekijöiden kannalta keskeistä AV-alan datassa on toisaalta heidän taloudellisten etujensa varmistaminen eli että he saavat kohtuullisen korvauksen työstään ja toisaalta heidän oman brändinsä rakentaminen eli katsojat ja mahdolliset yhteistyökumppanit tietävät kenen työtä he katsovat tai kuuntelevat.

Kun AV-alalla on dataa helpommin saatavilla, kumpikin tekijän tavoite on helpompi toteuttaa. Tekijän muut työt löytyvät helpommin sekä katsojan itse etsien että suosittelu-algoritmien suosittelemana. Jos teostieto on laadukasta, on käyttötietoja eri lähteistä helpompi yhdistellä ja siten tekijä saa myös luotettavampaa ja nopeampaa tietoa teostensa käytöstä. Tämä todennäköisesti helpottaa kohtuullisen korvauksen arviointia.

Tekijänoikeuden yhteishallinto-organisaatiot

Joissakin tapauksissa hyödyntäjän on tehotonta hankkia suoraan hyödyntämisoikeus kaikilta oikeudenomistajilta. Yksittäisen oikeudenomistajan intressi on pieni ja oikeudenomistajia on paljon. Näissä tekijänoikeuslaissa erikseen määritellyissä tapauksissa turvaudutaan Yhteishallinto-organisaatioihin, jotka antavat lisenssin kaikkien oikeudenomistajien puolesta.

Suomessa AV-alaa sivuavat keskeisimmät tekijänoikeusjärjestöt ovat tekijöitä edustava Kopiosto, tuottajia edustava Audiovisual Producers Finland APFI ja musiikin tekijöitä ja esittäjiä edustavat Teosto, Gramex ja heidän yhteistyötahonsa Musiikkiluvat.fi ja Nordisk Copyright Bureau.

Keskeinen osa yhteishallinto-organisaatioiden työtä on ylläpitää tietokantoja teoksista ja niiden oikeudenomistajista, selvittää teosten käyttöä ja yhdistää käyttötieto oikeudenomistajietoon. Laadukas ja nopeasti saatavilla oleva data tehostaa organisaation työtä, mahdollistaa automaatiota ja parantaa tili-tysten laatua.

Yhteishallinto-organisaatiot toimivat yhteistyössä saman alan ulkomaisten organisaatioiden kanssa. Tämä saattaa vaikuttaa suomalaisten yhteishallinto-organisaatioiden mahdollisuuksiin ja halukkuuteen ryhtyä kansallisiin yhteistyöprojekteihin.

Kanavat

Lineaarikanavat ovat edelleen merkittävin rahoittaja uudelle AV-tuotannolle. Kanavat haluavat lähtökohdaisesti mahdollisimman paljon katsojia. Kanavilla on usein tietty kohdeyleisö, joten keskeistä on saada katsojia tuosta kohdeyleisöstä.

Kanava tekee investoinnin ostaessaan sisältöä. Sisällön hankintaan liittyvän riskin hallinta on keskeistä varsinkin kaupallisille kanaville. Epävarmuutta aiheuttaa mainostajien ja katsojien reaktio ostettuun sisältöön. Jos ohjelma ei saa katsojia, ei se myöskään tuo mainoseuroja suunnitellusti. Tämän vuoksi muissa maissa testatut formaatit ovat tärkeitä kanaville.

Käytettävissä olevan datan lisääntyminen ja paremmat analyysipalvelut voivat antaa tietoa kuluttajien odotetusta reaktiosta tai peräti auttaa hallitsemaan noita reaktioita. Mainonnan kysynnän ja tavoitavuuden ennustaminen eri kohderyhmissä voi hyötyä laajemmasta datan saatavuudesta. Ostetun sisällön käyttämisessä on vapausasteita, esimerkiksi koska ja missä sisältö esitetään. Data helpottaa tätä optimointia.

Datan saatavuuden parantaminen edellyttää laajempaa tunnisteiden käyttöä. Laajempi tunnisteiden käyttö taas mahdollistaa ohjelmalogistiikan ja oikeuksien hallinnan prosessien tehostamista. Hyöty kanaville tulee siis myös prosessitehokkuuden parantumisesta.

Striimaajat

AV-sisältöjen kulutuksen kasvu on tapahtunut suurelta osin nimenomaan verkossa lähetettyjen sisältöjen ajamana. Tässä työssä tarkotamme striimaajilla toimijoita, jotka tarjoavat ammattimaisesti tuotettuja sisältöjä verkossa niin, että katsoja voi valita kirjastosta mitä ja milloin hän katsoo. Tarjolla oleva sisältö rahoitetaan joko mainoksilla, transaktio-pohjaisilla maksuilla ja/tai tilausmaksuilla.

Kuluttaja valitsee striimauspalvelun sen mukaan, mistä hän uskoo löytävänsä eniten häntä kiinnostavaa sisältöä. Tämän vuoksi palveluntarjoajat pyrkivät kokoamaan itselleen toisaalta laajan katalogin voidakseen tarjota jokaiselle jotakin, ja toisaalta paljon julkisuutta saavia uusia teoksia pärjätäkseen massamarkkinoilla. Laajan katalogin tarjoaminen saattaa antaa arvoa vanhemmallekin sisällölle. Jos datan jakaminen luo katsojille parempia mahdollisuuksia löytää sisältöjä, vaikuttaa se striimauksen kilpailutilanteeseen antaen myös niche-toimijoille tilaa.

Striimauspalvelulla on paljon dataa asiakkaidensa käyttäytymisestä. Datan jakaminen tuo kuitenkin striimauspalvelun tarjoajalle hyötyä, koska näin kehittyy vieläkin parempi ja laajempi data. Tämä taas mahdollistaa parempia työkaluja esimerkiksi asiakkaille tehtävien suosittelujen tekemiseen tai ohjelmien hankintaa edeltävien analyysien tekoon. Käyttäytymisdata olisi myös arvokasta tuottajille tulevien teosten kehittämisessä.

Teatterit ja festivaalit

Elokuvateatterit ovat erikoistuneet elokuvien ensimmäiseen ikkunaan. Heidän liiketoiminnassaan levittäjät ovat keskeisemmässä roolissa kuin tuottajat. Kuluttajan maksama hinta elokuvalipusta on korkeampi, joten kuluttaja odottaa saavansa enemmän vastinetta. Uutuuselokuvan lisäksi odotetaan huippuluokan valkokangasta, ääntä ja istuimia, joista syntyy yhdessä kokemus.

Kuluttajat ja mainostajat rahoittavat teatterin toiminnan. Filmikamari Ry edustaa suomalaisia elokuvateattereita sekä Suomessa toimivia elokuvien maahantuoja ja levittäjiä.

Festivaalien rooli on kaksinainen. Toisaalta ne ovat itsessään yleisöä palvelevia tapahtumia, joissa katsojat nauttivat elokuvateatterimaisesta kokemuksesta. Toisaalta ne ovat tuottajien näyteikkunoita uusille markkinoille. Agentit ja välittäjät etsivät uutta sisältöä festivaaleilta ja esimerkiksi USAn tai Ranskan markkinoille on vaikeaa päästä ilman festivaalipöhinää.

Paremman datan saatavuus parantaa ennustavaa analytiikkaa ja siten helpottaa ohjelmiston valintaa, tuottajan ikkunointia ja esittäjän sali / elokuva-mixin optimointia.

Riskirahoittajat

Eriyisesti elokuvissa ja KV-levitykseen tähtäävissä ohjelmissa saatetaan tarvita riskirahoittajia jakelijan ja Elokuvasäätiön lisäksi. Tämä johtuu mm. pidemmästä aikavälistä tuotannon aloittamisesta ensimmäisiin lipputuloihin sekä suuremmista kustannuksista per minuutti.

Riskirahoittajien toiminnassa keskeistä on mm. projektien valinta ja teoksen julkaisun jälkeen varmistaminen että kaikki heille kuuluvat tulot kotiutuvat heille. Projektien valinnassa on hyvä tietää projektin vastuullisten aiempien tuotantojen onnistuminen. Tulevaisuudessa tekoälyyn pohjautuvat analyysipalvelut voivat myös ennakoita käsikirjoituksen perusteella teoksen vastaanottoa. Molemmat hyötyvät paremmin saatavilla olevasta datasta. Jos teosten käytöstä kertyvää tietoa on laajasti saatavilla, on rahoittajan helpompaa luottaa ja varmistua, että he saavat kaikki heille kuuluvat osuudet lisensiointikorvauksista.

Agentit/levittäjät

Levittäjät ja agentit yhdistävät sisältöjen tuottajia ja ostajia. Tuottajat usein ulkoistavat AV-teoksien myynnin kotimarkkinoidensa ulkopuolelle siihen erikoistuneelle yhtiölle. Levittäjät/agentit arvioivat levityksen potentiaalin eri maantieteellisissä alueissa ja eri käyttöihin (festivaalit, elokuvateatterit, TV, striimaus) ja tuottavat näin lisäarvoa.

Alustatalous / digitaaliset jakelualustat saattavat syödä perinteisten agenttien elintilaa. Suurin syy alustojen menestykseen on näiden laaja data ja sen analysointi. Paremman datan saatavuudella levittäjät ja agentit pystyvät vielä paremmin arvioimaan kysyntää ja käymään läpi laajemmin tuotantoja löytääkseen niitä, jotka vastaavat kysyntään parhaiten. Tehokkaampi ja laajempi datan jakaminen myös tehostaa liiketoimintaprosesseja.

Tuottajakuluttajat (prosumers) ja nettisidosryhmät

Tuottajakuluttajat toimivat sosiaalisen median alustoissa, kuten YouTube. Nämä sisällöntuottajat pääsääntöisesti yhdistävät tuottajan ja tekijän roolit omista töissään. Heillä on tavallaan kaksi roolia tavallisen kuluttajan näkökulmasta. Toisaalta tuottajakuluttajat luovat sisältöä ja näin kilpailevat ammattimaisten tuotantoyhtiöiden kanssa kuluttajien ruutuajasta. Katsomme heitä kuitenkin ennen kaikkea kuluttajina, jotka käyttävät suojattuja teoksia osana omaa työtänsä ja toisaalta potentiaalisina liittolaisina tai alihankkijoina, jotka lisäävät teosten näkyvyyttä ja löydettävyyttä.

Tuottajakuluttajat eivät saisi vapaasti käyttää kaikkea löytämäänsä sisältöä toiminnassaan. Monet tämän ymmärtävätkin, mutta tarvittavan luvan hankkiminen koetaan hankalana. Samoin teosten, joiden tekijänoikeuden suoja on jo lakannut, identifiointi on hankalaa. Nettisidosryhmät tuottavat materiaalia AV-alasta ja AV-teoksista esimerkiksi IMDB ja Letterboxd-palveluihin ja some-kanaviin. Heille tämä on lähinnä harrastus.

Emme käsittele tässä työssä millaista dataa nettisidosryhmät käyttävät toimintansa optimointiin. Sen sijaan ammattimaisille tuottajille on kiinnostavaa miettiä millaista dataa syntyy heidän toiminnassaan. Analysoimalla suurta joukkoa näitä toimijoita voidaan analysoida yleisön reaktioita esimerkiksi elokuvan ensi-illan aikoihin. Samoin voidaan analysoida esimerkiksi IMDB:n arvosteluiden aikasarjaa.

Täydentävät palvelut

Esittelemme lopuksi muutaman tärkeän toimijan, joilla on keskeinen rooli AV-alan datan syntymisessä, jalostamisessa, siirtämisessä ja säilyttämisessä

Suomen elokuvasäätiö SES

Suomen elokuvasäätiö SES on merkittävä rahoittaja, mutta se myös tilastoi paitsi oman toimintansa vaikuttavuutta, myös elokuvien katselua Suomessa ja suomalaisten elokuvien käyttöä ulkomailla. Se jakaa dataa avoimesti.

Finnpanel

Finnpanel kerää dataa TV:n ja radion käytöstä. Heillä on 1000 satunnaisesti valittua TV-mittariperhettä, joiden kaikkea TV-sisältöjen käyttöä seurataan. Tätä täydennetään kanavien omista striimauspalveluista saatavilla tiedoilla. Yhteensä näissä perheissä on n. 2100 henkilöä. Tästä otoksesta mitattu TV-sisältöjen käyttö yleistetään tilastollisilla menetelmillä kattamaan koko väestön TV-sisältöjen käyttöä.

Finnpanelilla on keskeinen asema Suomessa kaikessa TV-sisältöjen katsomista ja tavoitavuutta koskevissa asioissa. Kanavat käyttävät ohjelman onnistumisen mittarina Finnpanelin laskemaa katsojalukua. Mainostajat ostavat peittoa tietyssä segmentissä, joka verifioidaan Finnpanelin katsojaluvuilla.

Lehdet ja verkkopalvelut

Kuluttajille halutaan saada tietoa tarjolla olevista teoksista. Lineaarikanavat jakavat tietoa suunnitelmastaan ohjelmakartasta lehdille, verkkopalveluille sekä suoraan kuluttajan laitteille. Standardi on nimeltään EPG.

Tämän lisäksi lähettäjäryitykset tarjoavat medialle lisätietoja venetsia-palvelun kautta.

Lehdet ja verkkopalvelut tarjoavat kuluttajille lineaarikanavien ohjelmatiedot, koska ne ovat kuluttajille mielenkiintoisia ja siten tuovat lisää lukijoita, joka taas tarkoittaa mainostuloja.

Kansallinen audiovisuaalinen instituutti KAVI

KAVI määrittelee oman tehtävänsä seuraavasti: "Kansallinen audiovisuaalinen instituutti (KAVI) on sekä muistiorganisaatio että valvontaviranomainen, jonka tehtävänä on säilyttää kotimaisen av-kulttuurin tuotteet sekä edistää audiovisuaalista kulttuuriperintöä ja mediakasvatusta."

Muistiorganisaationa sillä on tietoa AV-teoksista sekä halua parantaa arkistoidun lähetysvirran ja yksittäisten teosten käytettävyyttä esimerkiksi lisäämällä niistä saatavilla olevaa dataa. Lisäksi KAVI organisoi ja valvoo ikärajojen merkitsemistä kuvateoksiin.

AV-ekosysteemin data

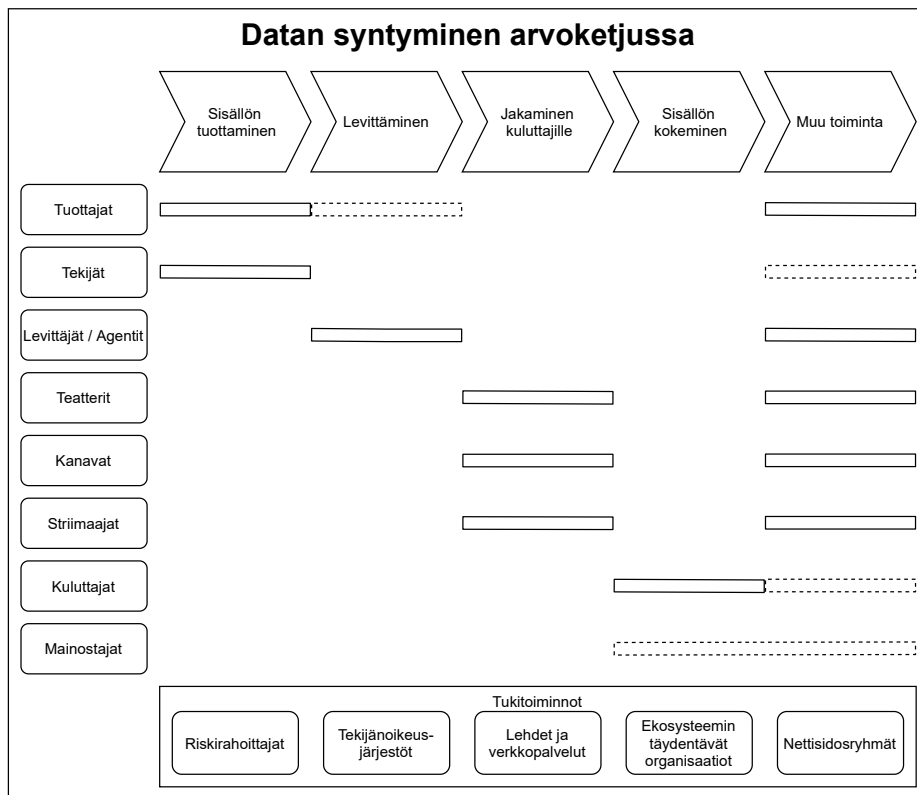
Tämän työn keskeisin fokus on data. Määrittelimme aiemmin AV-ekosysteemin koostuvan AV- sisällön tuottamisesta, sen tekoon ja jakeluun liittyvästä tekniikasta ja toimialaa säätelevästä regulaatiosta. Data-ekosysteemi voidaan määritellä olevan toimijoiden infrastruktuuri ja applikaatiot, jotka kokoavat dataa ja analysoivat siitä tietoa. Nämä yhdistämällä voimme muotoilla tämän työn fokuksen olevan AV-alan teoksista sekä niiden tuottamisesta, jakelusta ja kuluttamisesta syntyvän datan kokoamista ja jakamista, jotta siitä voidaan saada toimijoita hyödyttävää tietoa.

Datan synty

AV-alan arvoketjulle ei ole yhtä vakiintunutta kuvausta. Erityistä painoa kuvauksissa voidaan antaa sisällöntuotannolle, jonka vaiheita pilkotaan osiin. Erilaiset levitys- ja esitystavat vaikuttavat arvoketjuun: digitaalinen jakelu suoraan kuluttajille johtaa erilaiseen arvoketjuun kuin lineaari-TV.

Seuraavassa kuvassa on edellä esitellyt alan toimijat ja missä arvoketjun vaiheessa he osallistuvat datan tuotantoon. Käytämme laajaa datan määritelmää: esimerkiksi tekijän haastattelua voidaan käyttää tiedon louhintaan, joten se on dataa.

AV-alalla ei ole vielä Suomessa laajasti syntynyt yrityksiä, joiden ydin on datan jalostus. Niitä tulee varmasti syntymään joka portaalle, esimerkiksi tekoäly-sovelluksia tuovan prosessin tueksi, ikkunoiden ajoittamisen optimointiin tai ohjelmakartan analysointiin. Emme käsittele tätä tässä työssä enempää.



Kuva 1. Datan syntyminen arvoketjussa, kaikki datatyytit

Tuotantoyhtiö kehittää tarjolla olevista ideoista varsinaiseen tuotantoon sopivimman projektin. Tuotantoyhtiö siis kiinnittää tekijät, hankkii oikeudet kaikkiin tarvittaviin immateriaalioikeuksiin, hankkii rahoituksen (usein myymällä syntyvän teoksen esitysoikeuksia lähetysoikeuksille tai striimaajalle) ja hallitsee tuotantoprosessin valmisteluista varsinaiseen kuvaukseen ja kuvausten jälkitöihin.

Teoksen levittämisessä teos pyritään saamaan oikeaan aikaan oikeaan paikkaan. AV-teoksissa puhutaan usein ajallisista ikkunoista (esimerkiksi elokuvalevitys on 1. ikkuna, tallennemyynti 2. ikkuna ja lineaari-TV 3. ikkuna) ja eri maantieteellisistä markkinoista. Levittämisessä hallitaan näitä ikkunoita ja markkinoita.

Jakaminen tarkoittaa kaikkea TV-kanavilta, striimaajilta, teattereilta ja muilta toimijoilta vaadittavaa toimintaa, jotta teos saadaan kuluttajien nähtäville.

Kuluttaja katsoo teoksen ja kokee sen omalla tavallaan.

AV-teoksilla on usein pitkä elinkaari. Yleensä geneerisissä arvoketju-kuvissa nämä varsinaisen tuottamisen ja jakelun jälkeiset tehtävät määritellään asiakaspalveluksi. Kuitenkaan After sales /Asiakaspalvelu ei ole terminä AV-arvoketjussa relevantti. Olemme laajentaneet vaihetta muuksi toiminnaksi, joka kattaa mm. immateriaalioikeuksien hallinnan, kolmansien osapuolien tuottaman materiaalin ja sen hallinnan.

Datan tyypittely

Dataa voidaan tyypitellä monin tavoin. Aiemmin datan jatkokäytön takia oli tärkeää, että data oli määrämuotoista eli oikeassa datakentässä, oikeassa muodossa ja validoituna. Jatkuva kehitys on kuitenkin laajentanut ei-määrämuotoisen datan käyttökelpoisuutta; esimerkiksi Twitter-viesteistä voidaan louhia tietoa trendeistä.

Tässä työssä päähuomio on tekijänoikeusjärjestelmän tarvitsemalla tiedolla. Toiminta olisi tehokkainta, jos tieto olisi aina saatavilla määrämuodossa, jolloin teos, oikeudenomistajat eri käyttötapauksissa ja teoksen käyttö olisi tehokkaasti ja varmasti yhdistettävissä. Seuraavassa luokittelemme tiedon sen kohteen mukaan: tieto teoksesta, tieto oikeudenomistajista ja tieto teoksen käytöstä.

On tärkeää kuitenkin pitää mielessä, että datan käsittely muuttuu koko ajan. AV-alalla tämä voi vaikuttaa esimerkiksi teoksen identifiointiin, trendien ja yleisemmin kuluttajakäyttäytymisen ennustamiseen jakelejoilla, projektien valintaan tuottajilla ja suosittelualgoritmien kautta kuluttajien käyttäytymiseen. Datan käsittelyssä keskeisiä trendejä ovat suurempien datamassojen kerääminen ja prosessointi sekä ei-strukturoidun datan hyödyntäminen esimerkiksi tekstiä analysoimalla. Myös datasta syntyvään tietoon halutaan reagoida paljon nopeammin.

Tämä selvitys ei pureudu kaikkeen toimialan dataan. Tekijänoikeusjärjestelmän kannalta merkityksettömiä, mutta suomalaisen AV-toimialan kannalta tärkeitä ovat esimerkiksi tarkempi tieto (striimauspalveluiden) katselijoista. Kiinnostavaa dataa on myös mainonnan tai tuotteen kehittämiseksi tehty pienemmän koyleisön reaktioiden data. Katsojasegmenttien ja trendien analysointi kehittyy ja siinä käytettävä data ja syntyvä tieto ovat tärkeitä. AV-alan rahavirroista suuri osa tulee mainostajilta, jotka käyttävät mainostilaa niin kaupallisilla kanavilla, YouTubessa kuin myös esimerkiksi tuotesijoittelun (elokuvat, tv-sarjat) muodossa. Näin ollen mainostajien päätöksiä auttava tieto on toimialalle erityisen arvokasta.

Datan arvo ja merkitys

Tämä selvitys ja sen tulokset keskittyvät tekijänoikeusjärjestelmän kannalta keskeiseen tietoon. On syytä kuitenkin pitää mielessä, että tiedon merkitys kilpailuetuna lisääntyy koko ajan. Tästä syystä pohdimme hieman datan arvoa ja merkitystä.

Organisaatio voi lähestyä datan arvoa kahdelta suunnalta. Helpompi lähestyminen on miettiä mitä joku on valmis maksamaan datasta (ja miksi). Organisaation olisi kuitenkin syytä myös miettiä datan avulla tehtävien päätösten tuottamaa hyötyä. Jos kohderyhmän parempi tunteminen tuo esimerkiksi enemmän katsojia tuosta kohderyhmästä, voi datan arvoa arvioida tuota kautta saatavien mainoseurojen kasvuna ja mahdollisesti myös ohjelmahankinnan säästöinä.

Parempi data voi johtaa parempiin päätöksiin esimerkiksi markkinoinnissa, hinnoittelussa ketjun eri askeleilla ja operatiivisissa strategioissa, kuten esitysikkunoissa ja ohjelmien lähetysohjelmissa. Joissakin tapauksissa paremman datan tarjoaminen on edellytys, että teos ylipäätään löydetään; striimauspalvelussa jokin teos voi olla saatavilla, mutta jos sitä ei löydetä, ei tuolla saatavuudella ole merkitystä.

Tekijänoikeusjärjestelmän kannalta keskeinen data

Tekijänoikeusjärjestön täytyy kerätä tieto teoksista, joita järjestö edustaa, myöntää sisällön käyttäjille lisenssi saamiensa oikeuksien puitteissa, koota korvaukset käyttäjiltä ja jakaa nuo korvaukset oikeudenomistajille.

Tässä tehtävässä keskeisimmät tiedot ovat teostiedot, tiedot teosten oikeudenomistajista ja tiedot teosten käytöstä.

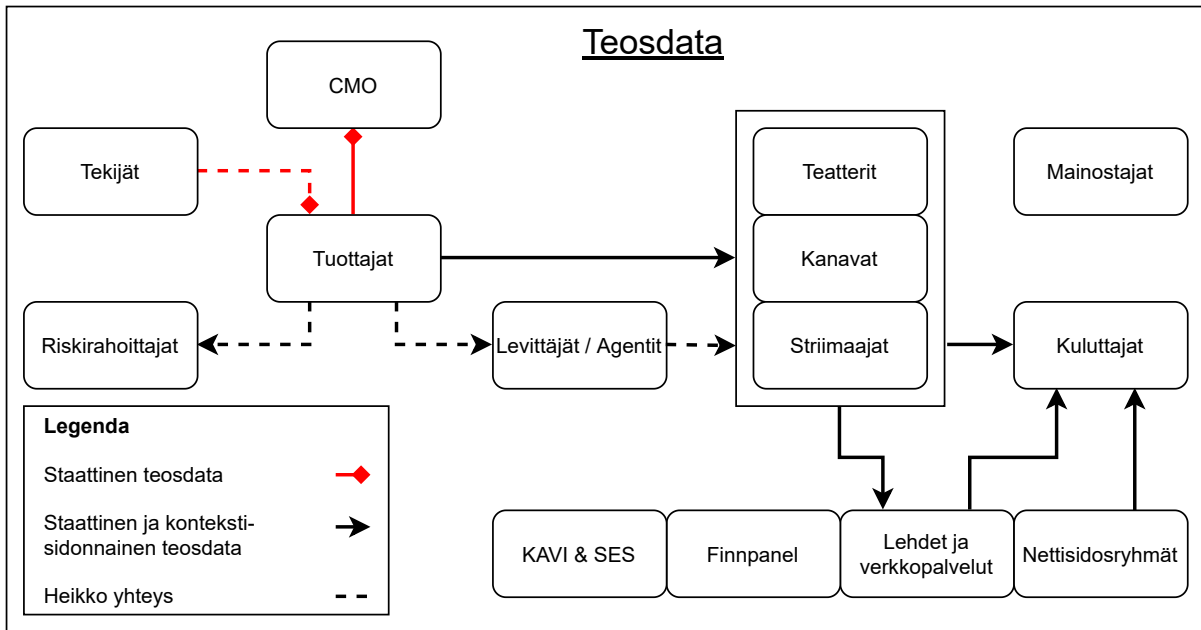
Teosdata

AV-teoksen luomisen aikana teoksesta syntyy dataa. Syntyvä data voidaan luokitella esimerkiksi tuotannon aikaiseksi metadatakksi, teoksen staattiseksi dataksi ja teoksen konteksti-sidonnaiseksi dataksi.

Tuotannon aikana syntyviin tiedostoihin käytettävät työkalut, kuten kamerat ja mikrofonit, liittävät syntyviin tiedostoihin automaattisesti lähinnä aika- ja tekniikkatietoa. Lisäksi manuaalisesti lisätään ainakin tunnistetieto, joka on Suomessa pääsääntöisesti lopputuloksena syntyvän teoksen nimi. Vaihtoehtona voisi olla uniikki tunniste, kuten kansainvälinen ISAN-koodi. Ranskassa on juuri käynnistynyt projekti, jossa teokselle haetaan ISAN-koodi jo tuotannon alkaessa, jotta eri yrityksissä tapahtuvaa jälkituotantoa pystytään paremmin hallitsemaan. Tuotannon aikainen metadata on mielenkiintoista, mutta siihen liittyvät toimintatavat eivät ole vakiintuneita eikä sitä nykyisellään liitetä millään tavalla lopulliseen teokseen. Emme käsittele sitä tässä selvityksessä enempää.

Kun teos valmistuu, siitä tarvitaan runsaasti tietoa. Tämän tiedon tarjoaa tuottaja osana teoksen julkaisua. Luonteeltaan tieto on staattista ja objektiivista, esimerkiksi teoksen kesto ilman mainoksia on 42 min. Keskitymme tähän tietoon.

Teoksesta syntyy myös kontekstisidonnaista, subjektiivista tietoa. Tällaista on esimerkiksi synopsis, genre ja arvostelut. Tällaista subjektiivista tietoa samasta asiasta voi olla useista lähteistä ja on tärkeää tietää, kuka tällaisen subjektiivisen tiedon on tuottanut. Koska tämä kontekstisidonnainen teostieto on keskeinen osa tuottajien ja lähettäjiä välistä kommunikaatiota ja siten hyvin potentiaalista materiaalia siirtymään tässä työssä hahmoteltavan portaalin kautta, on sille oma kappale erilaisia tekijänoikeuden kannalta keskeisiä dataja esittelevien kappaleiden jälkeen.



Kuva 2: Staatinnaisen ja konteksti-sidonnaisen teosdatan synty ja siirtyminen

Teosdatan keskeisin käyttö tekijänoikeusjärjestelmän kannalta liittyy teosten tunnistamiseen. Tämä tieto on objektiivista ja sen tulisi olla mahdollisimman määrämuotoista, jotta tunnistaminen voidaan luotetavasti tehdä. Esimerkkejä tästä datasta ovat alkuperäinen nimi, elokuvan/ jakson pituus, ohjaaja, näyttelijät ja tuotantoyhtiö. Jos on tarvetta yksilöidä myös teoksen (jakelu)versio, niin tarvitaan myös master-kopion tekniset tiedot. Tunnistamisongelman ratkaisisi globaalisti uniikki ID.

Tunnistamisen kannalta pitää pystyä erottamaan esimerkiksi sarjan kaudet toisistaan, sarjan jaksot toisistaan ja vielä eri versiot, kuten erilaiset tekniset formaatit tai kieliversiot toisistaan. Näin teosdata saattaa olla hierarkista eli esimerkiksi teoksen nimi on sama kaikille kausille, mutta kausien välillä kesto on eri. Kauden kestot ovat samat, mutta kullakin jaksolla voi olla omat näyttelijät. Jos tälle hierarkisuudelle ei ole selviä periaatteita, tiedon käyttö tunnistamiseen on hankalampaa.

Teosdatan lähde on tuottaja. Master-kopion valmistamisen yhteydessä tuotantoprosessin aikaiset tiedostoihin liitetyt metadatat hävitetään ja teoksen staatinnainen data kirjataan. Master-kopioon liitetään metadatan lähinnä teknistä tietoa ja teoksen nimi. Suurin osa tiedoista välitetään muita reittejä pitkin teoksen käyttäjille, kuten lähettäjäyrityksille ja/tai tekijänoikeuden yhteishallinto-organisaatioille.

Staatinnainen data siirtyy arvoketjussa teoksen kanssa tuottajalta jakelijalle ja edelleen kuluttajalle ja muille osapuolille muuttumattomana. Sen saatavuutta eri toimijoille kuitenkin voidaan rajoittaa eli esimerkiksi kuluttajien käyttöön menevä ohjelmatieto, EPG-tieto, ei sisällä kuin osan lähettäjän saamasta tiedosta. Jakelijat liittävät staatinnaiseen dataan konteksti-sidonnaista dataa, esimerkiksi synopsisen, välittäessään tietoja kuluttajille ja medialle.

Data oikeuksista

Koko AV-sisältöjen tuotanto nojaa tekijänoikeuksiin - teoksen tekijöillä on oikeus määrätä teoksen käytöstä ja saada käytöstä korvaus. Lähtökohtaisesti tekijänoikeus on henkilöillä, jotka ovat luoneet teoksen. Koska AV-teos on aina yhteistyön tulos, on tekijänoikeus silloin kaikilla näillä henkilöillä yhteisesti. Jotta teoksia voitaisiin järkevästi hyödyntää, täytyy tekijänoikeudet keskittää niin että käyttäjät voivat

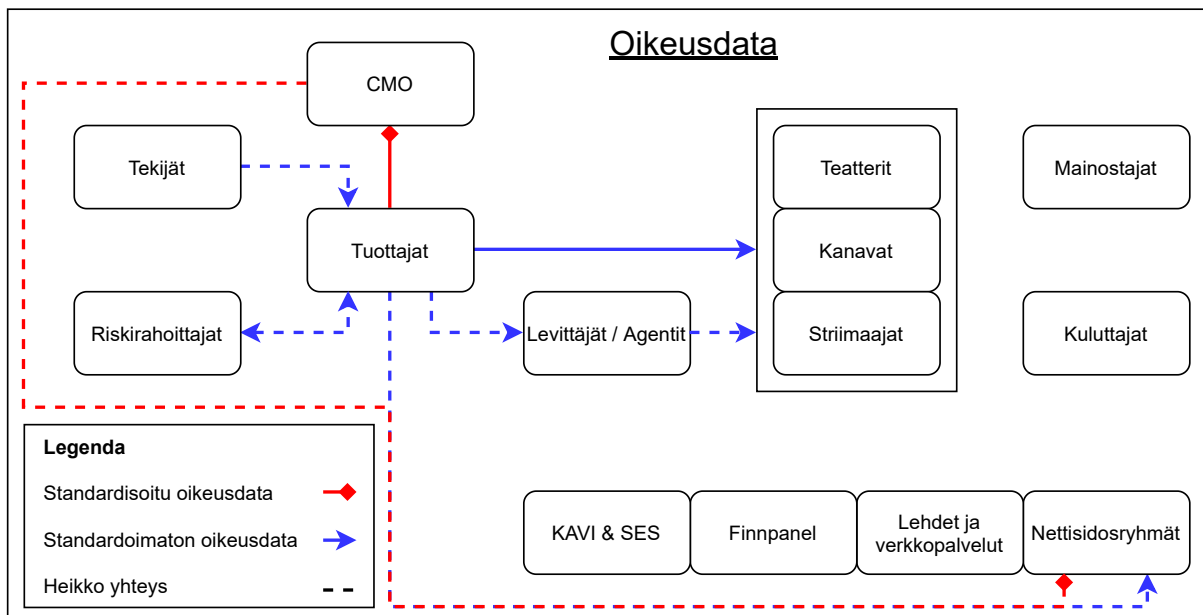
kohtuullisella työllä hankkia oikeuden käyttöä teosta.

Oikeus voidaan määritellä dataa ajatellen olevan "tila, jossa osapuoli saa tehdä jotakin teoksella ja joka seuraa lain määräyksestä tai sopimuksesta" (The LCC Rights Reference Model, s 4). Tekijänoikeudesta johdetut käyttöoikeudet ovat aina sidottuja tiettyyn tarkoitukseen ja kontekstiin eli käyttöoikeuteen vaikuttaa käyttötilanne; missä tehdään, mille tehdään, mitä tehdään ja kuka tekee. Esimerkiksi lähettäjäyhtiö XYZ on tällä sopimuksella hankkinut oikeuden esittää lineaari-tv-kanavillaan kahden vuoden aikana alkaen nn.nn.nnnn rajoittamattoman määrän lähetyksiä ohjelmasta zz. Tässä siis oikeutettu (yhtiö XYZ), peruste (sopimus), missä (kaikilla kanavillaan), mitä (esittää 2 v ajan), mitä (ohjelma zz). Tällä hetkellä suuri osa oikeuksista on kuvattu vapaatekstinä, joka edellyttää ihmisen tulkitsemaan onko käyttöoikeus olemassa tässä nimenomaisessa käyttötilanteessa.

AV-alan tekijänoikeusjärjestelmän kannalta keskeistä on yhteishallinto-organisaatioiden (Collective management organization, CMO) keräämä tekijänoikeustieto. Tämä on standardoitua, ennalta määrättyihin käyttötilanteisiin sovellettavaa lisensointia. Se on tarkoitettu lisensoimaan teosten käyttöä tilanteissa, joissa käytetään lukuisia teoksia, yhden teoksen saama korvaus on pieni ja teosten käyttäjien on vaikeaa tai mahdotonta saada suoraan sovittua teoksen käytöstä kaikkien oikeudenomistajien kanssa. Nämä tapaukset on säädetty erikseen tekijänoikeuslaissa. Musiikin puolella radiot maksavat myös korvaukset radiosoitosta yhteishallinto-organisaation kautta koska se on tehokasta, vaikka kyse ei olekaan tekijänoikeuslain mukaisesta yhteishallinnosta.

Mikromaksamiseen perustuvat jakelukanavat saattavat tuoda uusia standardoitua tekijänoikeustietoa tarvitsevia toimijoita. Samoin tekijänoikeuksien automaattinen valvonta digitaalisissa kanavissa saattaa tarvita standardimuotoista tekijänoikeustietoa.

AV-alalla teoksien käyttöoikeuksia myydään pääsääntöisesti erikseen neuvoteltavilla sopimuksilla, jolloin sopimuksen laajuus on erikseen neuvoteltavissa. Koska näitä sopimuksia tehdään kohtuullisen vähän, ei niiden standardointi ole prioriteetti, vaan laadinta ja tulkinta voidaan tehdä myös vapaamuotoisesti tekstistä.



Kuva 3: Tekijänoikeustiedon syntyminen ja siirtyminen

Tekijänoikeus syntyy luovan panoksensa antaneille tekijöille. Jotta oikeuksilla operointi on mahdollista, ne keskitetään tuottajalle. Tuottaja lisensoi teoksen käytön lähettäjälle tai ottaa agentin/levittäjän myymään teosta (erityisesti ulkomaille tai erityismarkkinoille). Tuottajalle keskitettyjen oikeuksien lisäksi yhteishallinto-organisaatioille voidaan luovuttaa oikeuksia lisensoida erikseen laissa määriteltyjä käyttötappauksia.

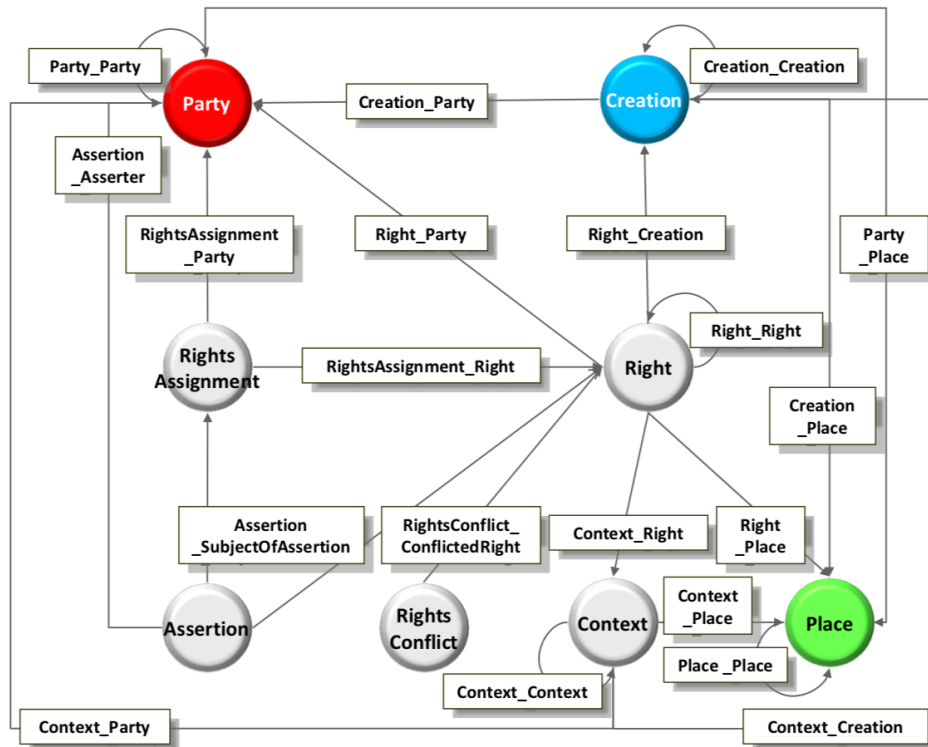
Kuvassa näkyvä yhteys riskirahoittajiin syntyy rahoittajien halusta varmistua, että teoksella on arvoa eli että se voidaan lisensoida käyttäjille. Samoin kuvassa on yhteys tuottajakuluttajille, joille lisensointi on tällä hetkellä AV-alalla olematonta, mutta se on mahdollinen kehityssuunta.

Suomessa tuottaja ilmoittaa tekijät Kopiostolle oman harkintansa ja asiantuntemuksensa mukaan. Tekijöiden ja tuottajien välisessä sopimuksessa ei ekspansiivisesti sovita, onko tämän tekijän panos niin suuri, että hänet ilmoitetaan Kopioston tekijä-rekistereihin, ja että hän näin saa yhteishallinto-organisaatiolta korvausta luovasta työstään. Kopioston vakiosopimus siirtää tekijän kaikkien olemassa olevien ja tulevien teosten oikeudet Kopiostolle.

AV-teoksissa käytettävän musiikin liittämiseen AV-teokseen tarvitaan (synkronointi)lisenssi, jonka voi hankkia suoraan oikeudenomistajalta tai yhteispohjoismaiselta NBC yhteishallinto-organisaatiolta. Joka tapauksessa AV-teoksen esittämisessä tarvitaan lisenssi myös musiikille. Näitä lisenssejä varten ilmoitetaan mitä musiikkiteoksia AV-teoksessa käytetään.

Tekijänoikeustiedon esittämiseen standardissa muodossa on abstrakti tietomalli Rights reference model. Sen on luonut Linked Content Coalition. Tuon mallin tietokohteina ovat seuraavan kuvan mukaiset. Lisätietoa mallista löytyy <http://www.linkedcontentcoalition.org>.

RRM Entity and Link Types



Kuva 4: Rights Reference Model - oikeuksien määrittelyn geneerinen muoto. Lainattu Linked Content Coalition LCC

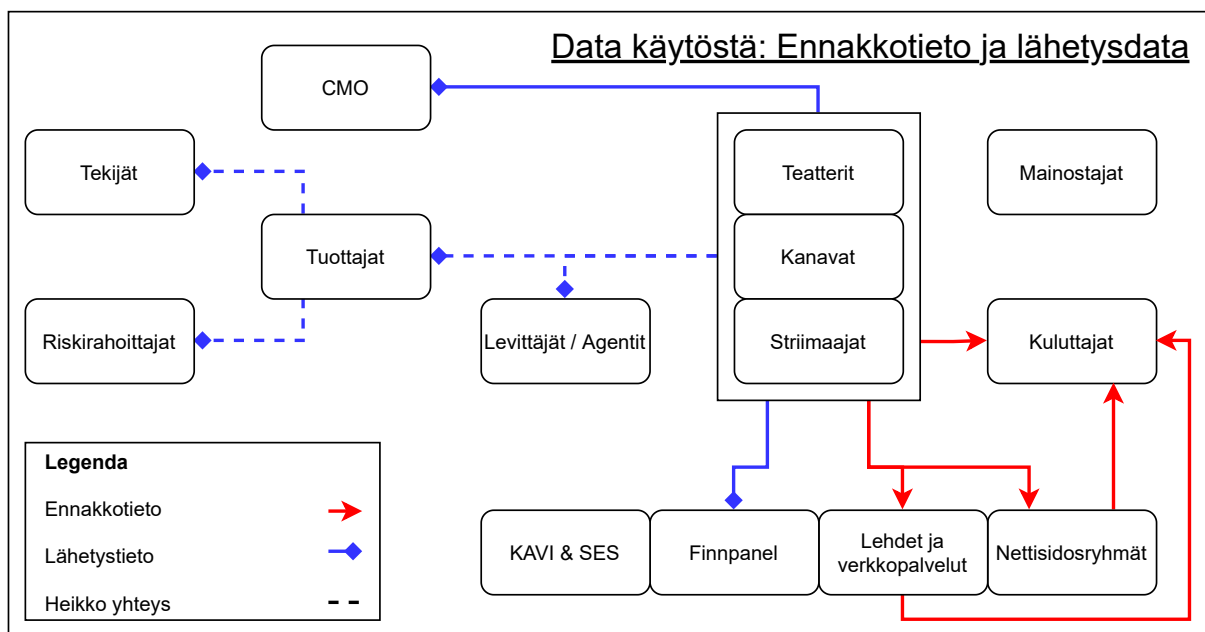
Data AV-teoksen käytöstä

Tieto AV-teoksen käytöstä on erilaista riippuen jakelutavasta. Käsittelemme ensin lineaari-TV:n ja sitten striimauksen.

Lineaari-TV:ssä teoksen käytöstä annetaan ensin ennakkotieto, jonka päätarkoitus on informoida kuluttajia siitä, mitä milläkin kanavalla esitetään ja koska.

Osana sisäistä prosessia lähetyksiä tekee ohjelmakartan. Lähetyksen jälkeen tämä päivittyä toteutuneeksi lähetystiedoksi, jonka perusteella esimerkiksi Finnpanel yhdistää TV-perheiden mitatun katsomisen varsinaisiin teoksiin ja yhteishallinto-organisaatiot tilittävät korvauksiaan.

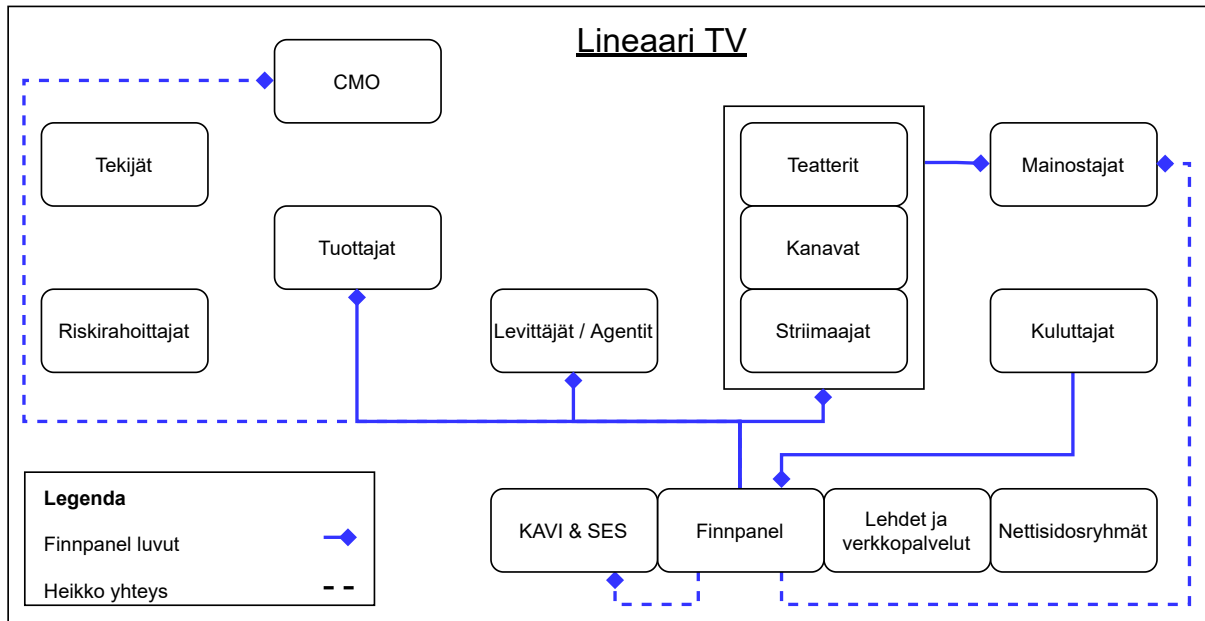
Lineaari-TV



Kuva 5: Ennakkotieto lähetettävistä ohjelmista ja lähetydata

Lähettäjäryitykset kertovat suunnitelluista lähetyksistä määrämuotoisella EPG-tiedolla, jota vapaasti jaetaan äly-tv-laitteille, medioille ja muille kiinnostuneille tahoille. Siinä on ohjelmien perustiedot eli ohjelman ja jakson nimi suomeksi ja alkuperäiskielellä, kausi, jakson numero ja synopsis. Suomessa on lehdistölle suunnattu, kaikkien kanavien yhteinen palvelu nimeltään Venetsia ja tästä syystä puhutaan myös Venetsia-tiedosta.

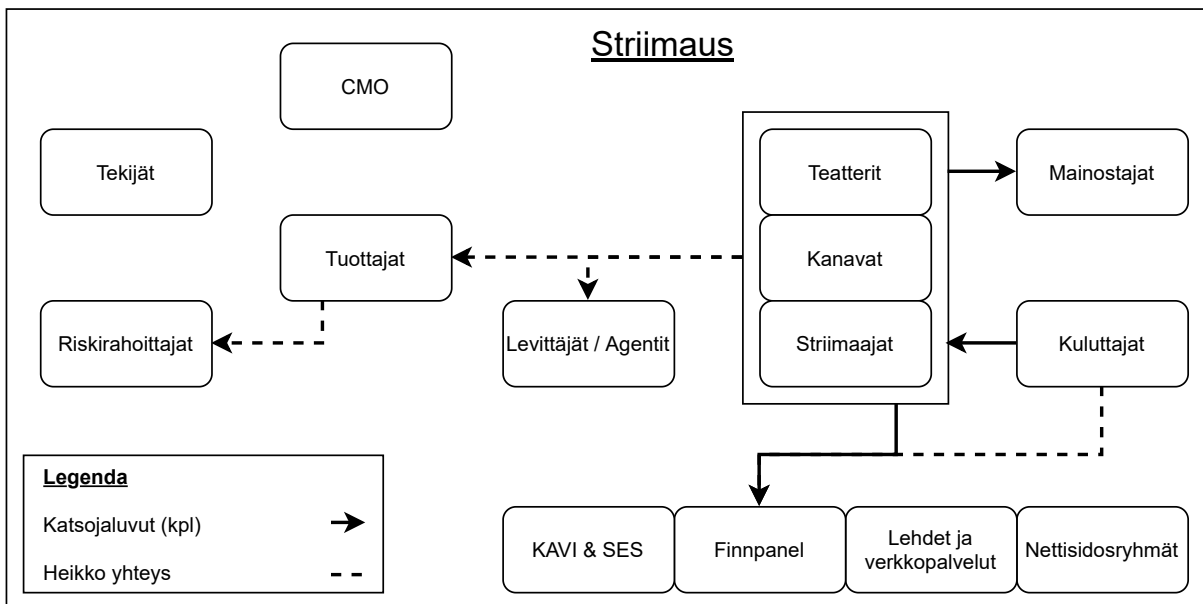
Lähettämistä seuraavana päivänä saatavilla on myös lähetystiedot, jossa on kanavan toteutunut lähetyks (sisältäen mainokset, trailerit jne). Lähettäjäryitykset antavat nämä tiedot haluamiensa tahojen käyttöön, esimerkiksi yhteishallinto-organisaatioille ja Finnpanelille.



Kuva 6: Lineaari TV:n katsojatietojen synty (Finnpanel) ja siirtyminen

Finnpanel kerää TV-perheistään kaiken TV:n käytön. Yhdistämällä tiedot perheiden TV:n käytöstä, perheenjäsenien demografiset tiedot ja Suomen väestön demografiset tiedot he voivat muodostaa luotettavan näkemyksen kunkin ohjelman keskimääräisestä katsojamäärästä katsojaryhmittäin. Tämä katsojamäärädata katsojaryhmittäin on julkisesti saatavilla, mutta tiedot TV-perheistä ja laskennan pohjana oleva toteutunut katsomistieto ei ole saatavilla.

Striimauspalvelut



Kuva 7: Katsojatietojen synty ja siirtyminen striimauspalveluissa

Striimauspalvelussa katsoja valitsee itse mitä hän katsoo ja koska. Striimauspalvelussa ei siis ole EPG-tietoon verrattavaa ennakkotietoa. Sen sijaan striimauspalvelulla on katalogi tarjolla olevista teoksista. Toi-

sin kuin lineaari-TV-puolella jonkin teoksen oleminen katalogissa ei automaattisesti tarkoita, että tieto tuosta teoksesta todellisuudessa voi vaikuttaa käyttäjien valintoihin siitä, mitä he katsovat.

Koska jokainen ohjelman käynnistys tapahtuu palveluntarjoajan palvelussa, on palveluntarjoajalla hyvin tarkka tieto käynnistyksistä ja myös toiston keskeyttämisen hetkestä. Palveluntarjoaja kykenee usein myös identifioimaan katsojan, joten hän pystyy toisaalta ryhmittelemään katsojia näiden katsomien teosten pohjalta ja mahdollisesti myös tietämiensä taustatietojen pohjalta. GDPR voi rajoittaa tätä mahdollisuutta. Tätä tietoa voidaan joissakin tilanteissa luovuttaa tuottajille.

Tarkka tieto lukuisten katsojien käyttäytymisestä tuottaa myös tietoa, jota voidaan hyödyntää suosittelualgoritmien kehittämisessä ja käyttämisessä.

Striimauksen erityistapaus, verkkotallennus, on yhteishallinto-organisaatioiden lisensoima. Muu striimaus kuin verkkotallennuspalvelu perustuu suoriin teoksen striimausoikeuksien ostoon eikä siitä synny mitään raportointitarvetta yhteishallinto-organisaatioille.

Konteksti-sidonnainen data

Konteksti-sidonnainen data muuttuu teoksen elinkaaren aikana. Ensimmäiset tällaiset tiedot laatii tuottaja; erimittaiset synopsikset, still-kuvat ja trailerit. Kuitenkin esimerkiksi jakelijat ja lähettäjärytykset muuttavat näitä tietoja paremmin maantieteelliseen alueeseen / asiakkaisiin sopivaksi. Samoin ajan kuluessa tämä tieto saattaa muuttua - esimerkiksi voitettut palkinnot ja saatu julkisuus halutaan tuoda mukaan esittelyteksteihin.

Tämän tiedon siirtäminen tuottajalta jakelijalle on osa teoksen toimitusta ja siten kriittistä tuottajalle. Kuitenkin tämä on suurelta osalta sähköpostien ja muiden ad hoc-toimintojen varassa. Prosessivarmuuden ja selkeyden kannalta olisi hyödyllistä tarjota automatisoitu työnkulku tuottajien ja jakelijoiden väliin samaan tapaan kuin kaupallisten TV-kanavien spotgate toimii mainoksille. Näin jakelija saisi sovituksessa muodossa paitsi staattisen teostiedon eli ohjelman ja jakson nimen, keston, ohjaajan jne., myös still-kuvat, trailerin, synopsiksen ja genren toimitettuna kanavan ilmoittamalle yhteyshenkilölle. Kaikki tämä tieto jäisi myös arkistoon, joten organisaatio olisi vähemmän haavoittuva yhden henkilön poistumiselle.

Erikseen on syytä mainita, että eri paikoissa käytettävät genret eivät ole pääsääntöisesti yhteensopivia tai toisistaan johdettavia.

Tuottaja-portaali -ratkaisu akuuttiin tarpeeseen

Datan jakamisen nykytila selvitettiin haastatteluilla. Käytössä on edelleen vanhat kahdenväliset datan siirtämisen ja jakamisen toimintatavat. Tuottajat välittävät vaaditut tiedot erikseen kullekin tekijänoikeuden yhteishallinto-organisaatiolle, kanavalle ja mahdollisille levittäjille. Yhteishallinto-organisaatiot eivät jaa tietoa keskenään esimerkiksi suomalaisten AV-teosten käytöstä ulkomailla. Yksiselitteisiä, laajasti käytössä olevia tunnisteita ei ole käytössä, joka vaikeuttaa tiedon siirron automatisointia.

Kehittymätön data-infrastruktuuri vaikeuttaa datan hyödyntämistä päätöksenteon tueksi. Datan merkityksen ja arvon kasvaminen ei hyödytä suomalaisia tuottajia, kanavia ja striimaajia. Näin suuret kansainväliset yritykset saavat merkittävää kilpailuetua. Monella muulla toimialalla on jo erikoistuneita data-jalostajia. Nämä erikoistuneet yritykset ovat tapa tasoittaa pelikenttää.

Nopeasti saatavia, kiistattomia etuja saadaan keskittelyllä datan syöttämisellä ja jakamisella. Samalla syntyy uniikki tunniste. Tämä toimintatapa myös mahdollistaa avoimemman datan jakamisen toimijoiden näin halutessa.

Keskitetty datan syöttö ja jakelu

Tuottajan odotetaan tuottavan teostiedon sekä lähettäjälle/striimaajalle että yhteishallinto-organisaatioille. Tietoa odotetaan eri vaiheessa projektia: ohjelmaidean myynti, musiikin liittäminen teokseen, teoksen perustietojen toimittaminen yhdessä master-nauhan kanssa, korvattavan luovan panoksen antaneiden tekijöiden ilmoittaminen yhteishallinto-organisaatiolle. Tällä hetkellä tieto välitetään jokaisen vastaanottajan omaan portaaliin tai sähköposteilla. Tätä prosessia voidaan keventää käyttämällä keskitettyä datan syöttöä ja jakelua.

Tämän selvityksen yhteydessä luotiin demo portaalista, jolla tuottaja voi keskitetysti syöttää datan eri vastaanottajille.

YHTEISET TIEDOT
KOPIOSTO
APFI
MUSIIKKI
KANAVAT

Täyttäjän tiedot

Etunimi Ville-Veikko	Sukunimi Virtala	Tuotantoyhtiö Filmi Suomi Oy	Päivämäärä 21.10.2021
Sähköpostiosoite ville-veikko.virtala@filmiy.fi	Puhelinnumero 045-121 2211	Katuosoite Teollisuuskatu 2	Postinumero 00510
Postitoimipaikka HELSINKI			

Teoksen tiedot

Jos teet jatkokautta, niin etsi hakutoiminnolla edellinen kausi ja kopioi se uuden kauden pohjaksi.
✕

Yksittäisteos
 Dubbaus suomeksi -ilmoitus

Sarja

Nimi

Tuotantomaa
Suomi

Kieli
Suomi

Kausi

Kesto ilman mainoksia

Tuottajan viite

Tilaajan viite

Alustava lähetyshyhtiö

Yksilöivät tunnisteet

Nimi muilla kielillä

KOPIOI UUDEKSI TEOKSEKSI

TALLENNA
LÄHETÄ

Kuva 8: Yleisten tietojen sivu portaalin demosta

Tunniste

Jotta eri lähteistä tulevia tietoja voidaan yhdistää, täytyy tietää mitkä tiedot koskevat samaa asiaa, esimerkiksi teosta tai käyttöä. Tarvitaan siis ID, joka mahdollistaa tämän. Paras ratkaisu on globaalisti uniikki ID. Esimerkki globaalista ID:stä on ISBN; jokaisella kirjalla on oma ID. Tämä mahdollisti Amazonin tarjoaman, jossa oli kaikki kirjat. Amazon ylläpitää verkkokauppaa ja välittää tuotteen ISBN-tiedon perusteella tilauksen fyysistä varastoa ylläpitävälle taholle. AV-alalla globaalit ID:t eivät vielä ole vallanneet alaa. Silloin täytyy luoda jokin muu uniikki ID. Mahdollisia lokaalisti uniikkeja ID:tä ovat mm. kunkin lähetyshyhtiön käyttämä Productcode ja yhteishallinto-organisaatioissa kussakin järjestön oma teos ID.

Teoksen uniikki ID yhdistää siis tietyn teoksen ja siitä tarjolla olevan tiedon. ID:n käyttäminen parantaa datan laatua ja käsittelyn tehokkuutta. ID:tä voidaan hyödyntää esimerkiksi erilaisten tuotantoyhtiön ulkopuolella tapahtuvien post production -töiden hallinnassa, teosten käyttötietojen ja oikeudenomistajatietojen yhdistämisessä yhteishallinto-organisaatioissa, lähetyisaikojen muuttamisesta aiheutuvien tallennusaikojen muutosten hallinnassa esimerkiksi verkkotallennuspalveluissa. ID mahdollistaa myös datan automatisoidun analysoinnin tuotantoyhtiöissä.

Lokaalisti uniikkeja ID:tä on jo käytössä lähetyshyhtiöissä ja striimaajissa ("productcode") ja yhteishallinto-organisaatioissa ("teos"). Näiden yhdistäminen täytyy kuitenkin tehdä käsin jokaisessa organisaatiossa. ID:tä ei yleisesti käytetä kommunikaatiossa toimijoiden välillä.

ISAN ja EIDR ovat eniten käytössä olevat globaalit ID:t AV-teoksille. Molemmat toimivat samalla tavalla eli tuottaja rekisteröi teoksensa tietylle numerolle, joka on sen jälkeen varattu vain tälle teokselle ja toimii teoksen uniikkina identifioijana aivan kuten henkilötunnus Suomessa. Koska ID:n täytyy identifioida teos kaikissa konteksteissa, täytyy eri kausilla, jaksoilla, versioilla ja teknisillä formaateilla olla oma ID.

Molemmat globaalit ID:t keräävät teoksen perustietoja kuten nimi, kesto ja tuotantovuosi. Suurin ero näiden välillä on niiden valitsema toimintamalli. ISAN perii palkkion per transaktio ja EIDR perustuu jäsenyyteen. EIDR siis on parempi suurille toimijoille ja ISAN harvoin tunnuksia tarvitseville toimijoille. Kumpikin pyrkii saamaan täyden linkkauksen toiseen järjestelmään.

Striimauksessa yksi mahdollinen ID on ISCC (international standard on content code), joka identifioi nimenomaisen verkossa liikkuvan teoskappaleen. Sen etu on, että ID muodostetaan itse teoskappaleesta eikä siis vaadi mitään erillistä rekisteröintiä.

Tuottaja-portaalin demo

Datan yhteiskäytön parantamisessa helppo askel eteenpäin on kerätä uusien teosten tiedot tuottajalta tehokkaasti, ja jakaa nuo tiedot tarvitsijoiden käyttöön luotettavasti. Samalla varmistuu, että syntyy tiedonvaihdossa tarvittava ID. Esimerkkinä toimintatavasta on Screenforce ry:n ylläpitämä Spotgate, jonka läpi kaikki suomalaisilla ilmaiseksi jaettavilla kanavilla näytettävät mainokset menevät.

Osana projektia tehtiin demo Spotgatea vastaavasta ohjelmien tietojen keruuseen sopivasta applikaatiosta. Applikaatiossa tiedon lähde, pääsääntöisesti tuottaja, syöttää kaiken tiedon yhteen paikkaan eli kehitettävään applikaatioon. Näin staattinen teostieto, kuten ohjelman nimi, kesto, ohjaaja jne., syötetään vain kerran. Samaan paikkaan syötetään eri vastaanottajille menevä tieto: musiikkitieto syötetään tänne ja se siirtyy automaattisesti musiikin yhteishallinto-organisaatioille, tuottajan laatimat synopsiset ja valokuvat siirtyvät esittävälle kanavalle ja niin edelleen.

Tässä selvityksessä toteutettiin ensimmäinen versio tuottajien syöttöikkunasta. Siinä on kerätty yhteen tuottajien raportointi Kopiostolle, Teostolle, APFille sekä ohjelman esittävälle lähetysyhtiölle.

Demo on esitelty omassa teknisessä dokumentissa, Demoon voi tutustua verkko-osoitteessa:

<https://data-alpha.vercel.app>

Demo on vasta osoitus ratkaisun toteutettavuudesta. Seuraavassa käydään läpi asioita, jotka täytyy vielä käydä läpi AV-alan ekosysteemin osapuolten kanssa. Keskeistä on vielä yhdessä varmistaa tuottajien, yhteishallinto-organisaatioiden ja kanavien kanssa lopullinen tarvittavien tietojen identifiointi ja tietokenttien yhdistely vastaanottajien järjestelmiin. Samalla keskustellaan myös ID-numeroiden tulevaisuudesta. Jotta yhteisestä tietojen keräämisestä olisi suurin mahdollinen hyöty, kannattaa myös toimintatapojen muuttamisesta ja tarvittavista koulutuksista keskustella yhteisesti. Ratkaisun turvallisuus ja kustannukset ovat tärkeitä ja ne täytyy keskustella yhdessä.

Tämän jälkeen tehdään varsinainen tietokanta sekä tietojen keräyksen vaatimat verkkolomakkeet ja mahdolliset rajapinnat.

Ratkaisu tehdään muutoksia kestäväenä modulaarisena ratkaisuna eli erikseen verkkolomakkeet, erikseen dokumenttitietokanta ja erikseen tietojen jakelun rajapinta. Ratkaisu on pilvipalvelu, jolloin käytettävissä ovat valmiit tietoturva- ym. palvelut.

Varsinaisen ratkaisun löytyminen vaatii vielä kaikkien osapuolien yhteistä työtä. Seuraavassa käsitellään edessä olevaa työtä.

Tarvittavan datan määrittely

Alustavassa demossa esitetyt tietokentät on kerätty yhteishallinto-organisaatioiden tiedonkeräyslomakkeista ja haastatteluista tuottajien kanssa. Niitä ei ole vielä tarkastettu tiedon vastaanottavien tahojen kanssa. Tällä hetkellä tietoa teoksista ja niiden käytöstä kerätään useassa organisaatiossa kunkin organisaation omiin tarpeisiin. Ennen datan keskitetyn keräilyn käyttöönottoa on varmistuttava, että jokainen organisaatio saa tarvitsemansa tiedon. Tämä tarkoittaa täytettävien tietokenttien tarkastamista ja niiden linkittämistä olemassa oleviin tietorakenteisiin, erilaisten ryhmittelyiden, kuten teostyyppien ja genren, tarkastamista ja sen varmistamista, että tarvittava ryhmittely pystytään tekemään kerätyillä tiedoilla.

On keskeistä huomata, että datan toimittamisen ajankohta ei ole sama kaikille vastaanottajille. Lähettäjäryitys haluaa ensimmäiset tiedot jo ennen tuotannon aloittamista, musiikkiluvat hankitaan jälkituotannon aikana, lähettäjä ja jakelija haluavat tietonsa master copyn valmistumisen aikoihin, ja tuottajien ja tekijöiden tekijänoikeusjärjestöt haluavat tiedot viimeistään lähetyksen tapahtuessa.

Toimintatapojen muutos ja sen vaatima kouluttaminen

Suurilla tuotantotaloilla on jo prosessit tietojen ilmoittamiseen yhteishallinto-organisaatioille. Uusi keskitetty ratkaisu tietojen ilmoittamiseen helpottaisi heidän toimintaansa. Kuitenkin eri vaiheissa tehtävät ilmoitukset eri vastaanottajille ovat mahdollisesti tuotantoyhtiössä eri henkilöiden vastuulla. Järjestelmän täytyy hallita tällaiset tilanteet.

Sen sijaan pienillä toimijoilla ei aina ole olemassa olevia prosesseja ja vastuita. Parhaimmillaan uusi keskitetty ratkaisu toisi selkeyttä kaikkiin vaadittaviin ilmoituksiin, erityisesti jos lähettäjäyhtiöt aktiivisesti sitä suosittavat sisältöjensä tuottajille ja tekevät tällaisen keskitetyn tietojen välittämisen helpoksi tuottajille.

Järjestelmän tulee tietenkin olla mahdollisimman selkeä käyttää ja syötettävien tietojen, esimerkiksi genren ja ohjelmatyyppien, mahdollisimman selkeitä ja hyvin kuvattuja. Tämän lisäksi käyttöönottoaiheessa täytyy järjestää riittävästi koulutuksia sekä käyttäjätuki. Käyttäjille on eduksi, jos he ymmärtävät miten ja missä heidän syöttämäänsä tietoa käytetään. Samoin tietojen käyttäjien on hyvä ymmärtää käyttämänsä tiedon syntyminen prosessi.

Ratkaisun kustannukset

Vuosituhanne alussa tekijänoikeusjärjestöt yrittivät keskittää datansa keruuta yhdessä. Tällöin hanke kuitenkin jäi mm. korkeiden kustannustensa vuoksi. Nykyiset pilvipohjaiset ratkaisut ovat merkittävästi edullisempia ja helpommin skaalautuvia.

Pyydetty alustava kustannusarvio portaalin luomiseen on n. 50 000 euroa. Kuukausikustannus verkkopalveluna toteutetusta applikaatiosta on alustavan laskelman mukaan alle 100 euroa kuukaudessa. Näiden kustannusten lisäksi tulee tietysti mukaan tulevien tahojen työpanos omissa organisaatioissaan sekä mahdolliset järjestelmäintegraatiot organisaation käyttämiin järjestelmiin.

Kustannuksia on käsitelty tarkemmin teknisessä kuvauksessa.

Turvallisuus

Yhteisen portaalin aikaansaamisen ja syntyvän datan tehokkaan hyödyntämisen yksi ehdoton edellytys on sen turvallisuus. Turvallisuudessa on monia ulottuvuuksia. Tietoturva ja muu tekninen turvallisuus on käsitelty demon teknisessä dokumentissa. GDPR-sääntely velvoittaa tarkastelemaan kaikkea tiedon keräämistä, tallentamista ja jakamista myös henkilötietojen käytön tarkoituksenmukaisuuden kannalta.

Hallinto

Tämän selvityksen pohjaksi otettiin SITRAn reilun datatalouden käsikirja. Sen lähtökohta on rohkaista samantyyppisistä asioista tietoa tuottavia, käyttäviä ja jalostavia tahoja yhdistämään tietonsa monenvälisen tietojärven avulla. Samalla kun sovitaan datan jaon periaatteista, niin sovitaan myös hallinnon periaatteista, kuten datan säilyttämisestä, datan elinkaaresta ja verkoston hallinnon periaatteista. Keskeinen osa hallinnon suunnittelua on myös pelisääntöjen sopiminen mahdolliselle dataverkoston laajentamiselle.

Tässä selvityksessä systemaattisesti tuodaan esiin keskitettyä ratkaisua, jossa data kerätään, talletetaan ja jaetaan yhdestä keskitetystä pisteestä. Tämä on ylläpidollisesti helpompaa. Hajautettu ratkaisu olisi mahdollinen toteuttaa käyttämällä API-rajapintoja näin haluttaessa.

Toimialan näkymät

AV-toimialalla on tapahtumassa merkittäviä muutoksia. Sisältöjen kulutusta siirtyy verkkoon. Alustapalvelut mahdollistavat sisällön kaupallistamisen uusilla tavoilla. Kuluttajien käyttäytymisestä kerätyllä datalla on arvoa. Seuraavassa esitellään muutamia näkymiä, jotka tulivat selvityksen aikana esille.

Tekijänoikeuksien luovutusten standardoimattomuus

Tekijänoikeuden luovutus on aina kontekstisidonnainen eli oikeus luovutetaan tiettyyn käyttöön, tietyksi ajaksi, tietylle maantieteelliselle alueelle.

Pääsääntöisesti yhteishallinto-organisaatioille luovutetut oikeudet ovat oikeuksia, joiden hankkiminen yksittäin suoraan oikeudenomistajilta olisi käyttäjille hankalaa ja/tai kallista verrattuna yksittäisen oikeuden arvoon. Nämä tapaukset ovat selvästi määriteltyjä ja siirrot ovat standardoituja.

Sen sijaan muiden oikeuksien luovutukset ovat usein standardoimattomia eli sopimuksissa luovutuksen laajuus kuvataan vapaana tekstinä. Tämä hankaloittaa automatisoituja oikeudentarkastuksia, jotka saataisivat olla tarpeen esimerkiksi mikromaksamisen yleistyessä tai youtubettajien hankkiessa oikeuksia omissa videoissaan käyttämiin videoklippeihin. Tällaisten liiketoimintamallien puuttuminen haittaa tulojen saamista pitkältä ajalta teoksen julkaisun jälkeen.

Standardoitut oikeudenluovutukset myös selkeyttäisivät oikeudenluovutusketjuja esimerkiksi ohjaajalta tuottajalle, tuottajalta levittäjälle ja levittäjältä lähettäjälle tai muulle käyttäjälle. Näin AV-alalla mahdollisesti vähenisi "kaikki oikeudet"-luovutukset.

Standardoitua tekijänoikeustietoa on AV-alalla lähinnä yhteishallinto-organisaatioilla niiden edustamien oikeuksien osalta. Yhteishallinto-organisaatioiden tieto ei ole avointa, joten on esimerkiksi vaikeaa selvittää, onko jokin teos vielä tekijänoikeuden suojaama, koska tietoa teoksen tekijöistä ja heidän kuolemastaan ei ole helposti saatavissa.

Oikeuden siirron sisältö voidaan systematisoida. Rights Reference Model määrittelee ulottuvuuksiksi mm. teoksen ("Kummeli-kultakuume", siirrettävän oikeuden ("Julkinen esittäminen"), paikan ("kaikki Suomen koulut"), osapuolet (tekijä, oikeudenomistaja, käyttäjä), kontekstin ("Opetuksessa sopimuskauden aikana").

Lohkoketjut

Lohkoketju-teknologialla voidaan varastoida tietoa hajautetusti ja luotettavasti. Se sopii siis myös tekijänoikeustiedon varastointiin. Varastoitava tieto voi olla standardoitua tekijänoikeustietoa eli onko taholla NN oikeus tehdä XX tai luovuttaa oikeus YY. Oikeuksien luovutus tietysti rekisteröidään lohkoketjuun.

Yksi erityissovellus oikeuksien luovutuksesta on non-fungible token (NFT), jossa verkossa oleva digitaalinen teos luovutetaan lopullisesti omistajalle. Tämä on jo käytössä musiikissa ja kuvissa, mutta AV-teosten suuri koko aiheuttaa ongelmia lohkoketjuverkossa.

Data-intensiteetin kasvu ja avoin data

Sisältö on edelleen kuningas AV-alalla. Kuitenkin datan hyödyntäminen alalla on suuressa muutoksessa arvoketjun jokaisessa askelmassa. Tuotannossa luovaa prosessia voidaan nopeuttaa hieromalla käsikirjoitusta tekoälyn avulla tai arkistosta voidaan paremmin löytää käyttökelpoisia leikkeitä. Tuotantoprojektien valinta tuotantoyhtiössä ja ohjelmien osto ja ohjelmakartan laadinta voivat hyötyä trendien, segmenttien ja niiden käyttäytymisen kattavammasta analysoinnista. Kuluttajat löytävät verkkohauistaan uusia teoksia ja samalla jättävät analysoitavaa dataa. Yhteishallinto-organisaatiot tehostavat toimintaansa ja parantavat tilitystensä tarkkuutta. Tuottajat ja kanavat voivat analysoida teosten menestymistä luotettavammin ja tehokkaammin.

Muutkin toimialat ovat kokeneet data-intensiteetin kasvun; datalla on arvoa. Tuo arvo perustuu datan mahdollistamaan parempaan päätöksentekoon. Google yhdistelee käyttäjän verkkohakujen tietoja ja näyttää sen ansiosta kohdistettuja mainoksia. Sähköyhtiö ennustaa historia-aikasarjojen ja sääennusteiden perusteella sähkönkulutusta tulevaisuudessa ja optimoi sähkönhankintaa sen mukaan.

AV-alalla syntyvää, tekijänoikeusjärjestelmälle olennaista tietoa on esitelty omassa kappaleessaan sivulla 14. Keskeisimmät datakokonaisuudet tekijänoikeusjärjestelmälle ovat tieto teoksista, tieto teosten käytöstä / katsomisesta ja tieto teosten oikeuksien omistajuudesta. Dataa kuitenkin hyödynnetään koko ajan uusista asioista ja lähteistä. Jo nyt on käynnissä pilotit datan louhinnasta esimerkiksi kuvailevista tekstityksistä ja käsikirjoituksista. Verkkopalveluiden eksaktit käyttötiedot ovat tällä hetkellä ainoastaan palveluntarjoajien käytössä, jotka hyödyntävät tietoja lähinnä suosittelu-algoritmeissa.

Datan laatu ja saatavuus vaikuttavat syntyvän tiedon käyttökelpoisuuteen ja siten nopeuteen, jolla pääsemme suuremman lisäarvon ja paremman kannattavuuden tietotalouteen. Muilla toimialoilla kolmannet osapuolet tuottavat data-analytiikkapalveluita esimerkiksi syvällisemmästä käyttäjätiedosta. Datan jakaminen lisää luonnollisesti läpinäkyvyyttä ja luottamusta dataan, datan prosessoihin ja laajasti alalla toimijoihin. Kuluttajat haluavat lisäarvoa tiedosta. Datan avoimuus lisää kaikkien luottamusta ekosysteemin toiminnan reiluuteen ja hallittavuuteen.

Suomalaiset alan toimijat ovat maailman mitassa pieniä. Jos jokainen suomalainen AV-alan yritys säilyttää oman datansa siilossaan, ei laadukasta dataa Suomesta suomalaisille toimijoille ole saatavilla. Tämä lisää suurten alusta- ja sisältöjättien valtaa ja pienemmät yritykset ovat merkittäväällä takamatkalla. Muilla aloilla datan jakamiseen on rohkaistu mm. SITRAn reilu datatalous-hankkeilla.

Laadukkaan datan saatavuus mahdollistaa dataa analysoivien yritysten synnyn. Jo nyt on syntymässä yrityksiä, jotka tekoälyn avulla pyrkivät esimerkiksi analysoimaan katsojien haluaman sisällön muuttumista (esim. Parrot Analytics) tai nopeuttamaan luovaa prosessia (esim. Largo). Jos dataa jaetaan, voi syntyä palveluyrityksiä datan ympärille. Data-analyysi on suuremmille toimijoille mahdollinen omien datavarastojen ja analyysikyvykkyyden ansiosta, mutta mahdollisuuksien saaminen maailman mittakaavassa pienempien toimijoiden ulottuville edellyttää aktiivisia päätöksiä toimialalla.

Valinta/suosittelualgoritmien kehittäminen

Toimialalla tutkitaan paljon kuluttajien käyttäytymistä. Dataa muodostuu esimerkiksi elokuvien esittämisestä testileisölle, kyselytutkimuksista median käytöstä, miten mainoskampanjat on huomattu ja kohderyhmää saavutettu sekä mitä ohjelmaa on käynnistetty verkkopalveluissa.

Näin saatavaa tietoa voidaan käyttää erilaisten optimointialgoritmien kehittämiseen. Tutuin sovellus on YouTuben suosittelu-algoritmi, joka seuraa yksittäisen käyttäjän verkkokäyttäytymistä ja vertaa sitä muiden käyttäjien käyttäytymiseen. Tämän perusteella se voi suositella osuvia ohjelmia. Kuluttajien verkkopalvelut suosittelevat samalla tavalla. Staattisia VOD-sisältöjä on helpompi suositella, lineaari-TV:n ohjelmavirrasta suosittelu on hankalampaa. Lisäksi lineaarivirrasta suosittelu ei tuo verkkopalvelun tarjoajalle samalla tavalla lisämyyntiä.

Verkkopalveluista saadaan transaktiotietoa. Tietoa on merkittävästi enemmän kuin kyselyillä tai Finnpanel-tyyppisillä otoksilla on saatavissa ja se on yksityiskohtaisempaa. Tätä tietoa voidaan käyttää kuluttajakohteisena aikasarjadataa. Yhdistämällä lukuisten käyttäjien tietoja saadaan luotua ryhmiä ja kohdistettua toimia kuten mainontaa tai ohjelmasuosituksia tarkemmin. Vertaamalla käyttäjän katseluhistoriaa muiden käyttäjien historiaan, voidaan tehdä suosituksia muiden käyttäjien historian perusteella.

AV-alalla on runsaasti ilmeisiä optimointikohteita. Mainonnan tyyli halutaan sopeuttaa valittuun välineeseen ja kohdeyleisöön, joten parempi tieto kohderyhmän luonteesta erityisesti valitussa välineessä on arvokasta. Ohjelmien valinta kanavalle perustuu ammattilaisiin, mutta sitä voisi täydentää analyysitiedolla esimerkiksi ulkomailta, edellisiltä vuosilta ja verkkokeskusteluista louhitusta tiedosta.

Suomessa erittäin keskeistä on myös Finnpanelin tekemä datan jalostus. He keräävät otoksen AV-sisältöjen käytöstä ja laajentavat tuon otoksen tilastollisiin työkaluin koko Suomen katsojia kattavaksi. Tuota tietoa käytetään mm. teosten tavoitteiden asettamiseen ja mainonnan kattavuuden validointiin.

Jos AV-ala seuraa muiden toimialojen kehitystä, alkaa nykyisen kuvaavan mallintamisen jatkoksi syntyämään ennakoivaa ja ohjaavaa mallintamista / analyysia. Lisääntyneen datan hyödyntäminen kuitenkin edellyttää erikoistunutta osaamista datan yhdistelystä ja datan louhinnasta esimerkiksi tekoälyn avulla.

Lainsäädännön vaikutukset

Käynnissä oleva DSM-direktiivin implementointi saattaa edistää tiedon jakamista eri toimijoiden kesken. Keskeisimmät artikkelit tämän selvityksen kannalta ovat art. 17 (verkkoalustan vastuu tekijänoikeudesta) ja art. 19 (tekijän oikeus saada tietoa teoksensa hyödyntämisestä).

Artikla 17 määrää, että verkkoalusta on vastuussa alustan käyttäjien palveluun siirtämästä sisällöstä, jos se loukkaa tekijänoikeutta. Tämä tulee johtamaan AV-teoksissa jonkinlaiseen ennakkotarkistukseen, jotta alusta osaa pysäyttää palveluun ladattavat, tekijänoikeutta rikkovat sisällöt jo ennen kuin ne päätyvät yleisön saataville.

Suurimmilla toimijoilla on edellytykset rakentaa yhteydet kaikkiin alustoihin, jotta he voivat toimittaa tarvittavat tiedot eston perustaksi. Sen sijaan pienemmillä toimijoilla ei ole voimavaroja tähän, joten joko tarvitaan yhteinen datan keräämisen ja jakamisen alusta tätä varten tai alalle syntyy kaupallisia valvontatietoja välittäviä yrityksiä.

Artikla 19 määrää, että tekijöillä on oikeus saada tietoa heidän teostensa hyödyntämisestä, jotta he voivat varmistua, että heidän saamansa korvaus tekijänoikeuden luovutuksesta on kohtuullinen. Kaikki AV-teokset ovat aina työryhmän tuloksia. Jotta artiklan tavoite toteutuu, täytyy olla olemassa standardoitu data, jossa on rekisteröity kaikki tietyn teoksen syntyyn itsenäisen luovan panoksen antaneet henkilöt. Tarkastusmahdollisuuden vuoksi olisi hyvä, että nämä rekisteröinnit olisivat julkisia.

Toisaalta täytyy olla rekisteri kyseisen teoksen oikeuksien luovutuksista ja niistä saaduista korvauksista. Tämä tieto on hyvin arkaluonteista kilpailutietoa. Tästä syystä artiklan ulottaminen AV-teoksiin on kiistanalaista.

Jotta AV-alalla raportointivelvollisuus saadaan täytettyä, täytyy yritysten kerätä standardimuotoon teki-jätieto, yhdistää tuo tieto teoksittain tuloihin ja tehdä raportointi. Se tarkoittaa joko tietojärjestelmäin-vestointeja, kaupallista palveluntarjoajaa tai avoimempaa tietojen jakamista toimialan sisällä, tietenkin GDPR huomioon ottaen.

Kansainväliset toimijat

Tutkimuksen aikana tuli monta kertaa esille, että Suomen AV-ala on monin tavoin sidoksissa kansainvä-liseen AV-alaan. Tuon vaikutuksen parempi ymmärtäminen vaatisi perusteellisempaa tutkimusta. Esil-le tuli kuitenkin esimerkiksi seuraavia havaintoja. Monet suuret tuottajat ovat kansainvälisten yhtiöiden tytäryhtiöitä. Se vaikuttaa esimerkiksi tietojärjestelmäkehitykseen sekä oikeuksien hankkimisen ja luo-vutusten periaatteisiin ja standardisopimuksiin.

Isojen striimauspalveluiden vaikutus suomalaiseen AV-alaan on vielä näkemättä. Toisaalta tekijät ja tuot-tajat hyötyvät, koska heidän työnsä on enemmän ostajia. Toisaalta kiristynyt kilpailu voi ajaa vanhoja lineaari-TV- lähettäjäyhtiöitä hankaluuksiin. Datan harkittu jako alan sisällä voi vaikuttaa kilpailutilantee-seen nimenomaan suomalaisen AV-alan eduksi.

Datan jako voi myös auttaa suomalaisia sisältöjä ulkomailla. Kun suomalaisesta sisällöstä on enemmän tietoa saatavilla, voivat kuluttajat helpommin löytää sitä. Määrätietoisella työllä voidaan mahdollisesti myös vaikuttaa suosittelualgoritmien toimintaan ulkomailla.

Tuottajakuluttajat

Tuottajakuluttajat (prosumers), esimerkiksi YouTubeen sisältöjä tekevät toimijat, ovat uusi erityinen ryh-mä toimialalla. He tuottavat sisältöä kuten AV-alan perinteiset tuottajat. Sekä sisällön tekemisen tekniikka että omien sisältöjen kaupallistaminen kehittyvät nopeasti ja lisäävät tämän tyyppistä toimintaa. Kuiten-kin näyttää siltä, että tätä YouTube-sisältöä kulutetaan ainakin vielä tällä hetkellä eri tavalla kuin ammat-timaisesti tuotettua sisältöä. Perinteiset tuottajat ja kanavat eivät voi kuvitella toimivansa irrallaan tuot-tajakuluttajista ja heidän sisältöjään jakavista alustoista.

Toisaalta tuottajakuluttajat ovat potentiaalinen asiakasryhmä: osa heistä käyttää teoksissaan tekijänoike-udella suojattua materiaalia, ja he olisivat valmiita suorittamaan siitä pienen korvauksen. Samoin nämä tuottajakuluttajat, some-julkikkiset, ovat entistä tärkeämpi markkinointikanava ja voivat olla perinteis-ten AV-alan toimijoiden yhteistyökumppaneita.

Sisältöjen laitton käyttö, piratismi ja mikromaksaminen

Suomalainen AV-ala menettää miljoonia vuodessa piratismiin vuoksi. Nordic Piracy -raportin mukaan Suomessa katsottiin vuonna 2020 laittomista lähteistä noin 6 miljoonaa kertaa suomalainen elokuva ja 12 miljoonaa kertaa suomalaisen TV-sarjan jakso.

Laitonta käyttöä tapahtuu monin tavoin. Olemassa olevan teoksen liittäminen osaksi uutta teosta on lähtökohtaisesti laitonta, vaikka tämä mielletään helposti laajennetuksi jokamiehen oikeudeksi. Käyttä-jä voi tällä hetkellä saada tuloja myös leikkaamalla paloja olemassa olevista teoksista ja laittamalla nii-

tä tarjolle verkkoalustalle. Lisäksi tarjolla on laillisia jakelutapoja imitoivia, mutta laittoman materiaalin päälle rakennettuja striimauspalveluita.

Jos AV-teoksilla on ID ja teoksesta on olemassa omistajatieto, voidaan tuottajakuluttajille tarjota lisenssin ostamista mikrotransaktiolla ja katsojille mahdollisuutta tarkastaa sisällön laillisuus ja ilmiantaa laitton sisältö. Koska verkossa palveluita tarjotaan miljoonille ihmisille, näistä saatavat pienetkin kertamaksut ovat helposti merkittäviä rahavirtoja.

Legacy tuotteet

Yhteishallinto-organisaatioiden tietokannoissa on kansainvälisesti jo olemassa miljoonia teoksia. Monia noista teoksista ei tulla käyttämään uudelleen lainkaan tai ainakaan usein lineaari-TV-puolella. Niihin ei siis liity suuria taloudellisia intressejä. Kuitenkin jotta tunnisteesta olisi maksimaalinen hyöty, olisi kaikille teoksille saatava tunniste. Näiden olemassa olevien teosten käsittely tulee olemaan yksi haaste eteenpäin pääsemisessä.

Suomessa yksi ratkaisu voisi olla kääntäminen kanavien puoleen. Kanavilla on "productionID", joka yksilöi jokaisen kanavalla lähetetyn teoksen. Joillakin yhteishallinto-organisaatioilla on jo "productionID" omista järjestelmissään.

Tästä huolimatta olemassa olevien teosten rekisteröinti vaatii rahaa. Tuottajalla ei enää ole intressiä tähän sijoittaa - teos on tehty ja myyty levitykseen eikä sen enää odoteta tuottavan. Kanavilla on oma ID ja heillä ei ole sen laajempia tarpeita. Yhteishallinto-organisaatioilla ei ole varaa käydä läpi koko teoskantaansa.

Jos tunnisteita ei saada, niin prosessien automatisointi on vaikeampaa. Samoin analytiikassa saattaa tulla virheellisiä johtopäätöksiä, jos osaa raakadatasta ei pystytä identifioimaan. Analyyseissä pystytään usein kuitenkin huomioimaan reikäinen data ja tuottamaan lisäarvoa datan ongelmista huolimatta.

Maininnan arvoisena erityistapauksena on niiden teosten listaus, joiden tekijänoikeus on jo päättynyt. Nykyisellään tuon tiedon saaminen AV-teoksista on hankalaa, ellei mahdotonta.

Yksittäisen ohjelman osat

TV-ohjelmat saattavat sisältää useamman oikeudenomistajan töitä eli esimerkiksi Pikku Kakkonen sisältää paitsi YLE:n itse tuottamaa juontoa ja musiikkia, myös ulkoa ostettuja ohjelmia. Tällaiset ohjelmat merkitään vaihtelevasti eri tietolähteissä joko osinaan tai yhtenä makasiiniohjelmana. Näiden ohjelmien merkintään olisi hyvä saada yhdenmukainen toimintatapa.

Arkistokäyttö lisääntyy eli aiemmin kuvattua materiaalia käytetään uudestaan. Tämä edellyttää tietenkin tietoa siitä, mitä ohjelma sisältää. Ratkaisuna on yritetty esimerkiksi runsaampaa metadataa, joka esimerkiksi kertoisi että "3:02 presidentti Kekkonen kävelee Tamminiemen puistossa". Myös kuvailutulkkausien hyödyntämistä asiasanoittamisessa on kokeiltu.

Projektin jatkoaskelmat ja rajoitteet

Selvityksen jälkeen tarkoitus on kehittää portaali, jolla tuottajat voivat tehostaa ja hallita paremmin teokseen liittyvää tiedonsiirtoa eri tahoille, kuten kaikille tekijänoikeusjärjestöille.

Selvitys ei ole tyhjentävästi kuvannut kaikkia tuottajien yhteistyötahoja. Projektin laajuus voisi siis helposti kasvaa kattamaan esimerkiksi elokuvasäätien kanssa kommunikoinin, ikärajojen määrittämisen tai post production -alihankkijaverkon hallinnan. Tärkeintä on kuitenkin päästä liikkeelle fokusoidulla tavoitteella. Tarkoitus on tehdä portaali modulaariseksi, jolloin uusien toiminnallisuuksien lisääminen on helppoa.

Tulevaisuudessa keskeistä on saavuttaa riittävä kattavuus tietojen vastaanottajissa (tekijänoikeusjärjestöt ja kanavat) sekä saada kanavien ja suoratoistopalveluiden hyväksyntä. Jos kanavat alkavat vaatimaan portaalin käyttöä, se nopeuttaa ratkaisevasti tuottajien liittymistä mukaan.

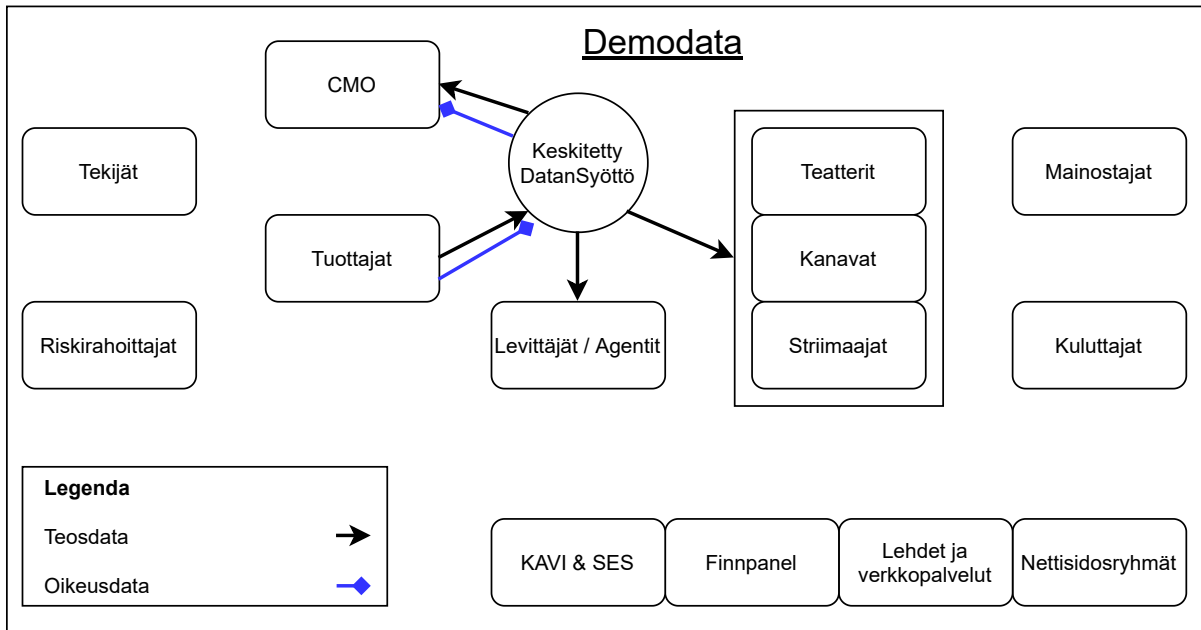
Datan hyödyntäminen on jo muodostanut suurille striimaajille kilpailuetua. Ainoa keino vähentää tuota kilpailuetua on lisätä datan jakamista AV-alalla Suomessa. Se vaatii tavoitteellista toimintaa. Seuraavassa projektissa voidaan vain luoda pohjaa tiedon jakamiselle; päätös tiedon jakamisesta on periaatteellinen jokaiselle osapuolelle, joten sitä ei voi edistää projektein. Laajemman tiedon jakamisen tulisi kuitenkin olla tavoitteena.

Yhteenveto

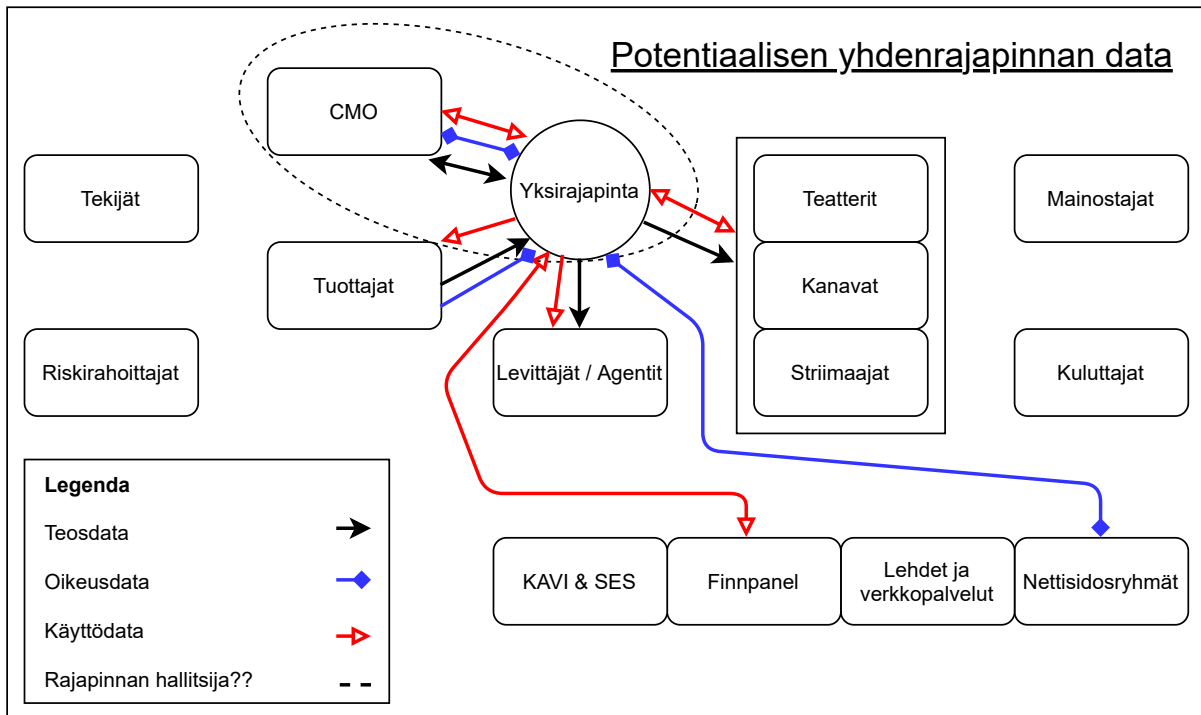
Datan keskitetty kerääminen, uniikin ID:n käyttöönotto, tiedon tehokkaampi siirto organisaatiolta toiselle ja datan avoimempi jakaminen hyödyttää koko toimialaa. Prosessin kustannukset laskevat, laatu paranee, datasta jalostettavasta tiedosta saadaan parempaa ja se on käytettävissä nopeammin.

Tämän selvityksen yhteydessä tehty demo osoittaa, että keskitetty tiedon kerääminen on tehtävissä ja se parantaa tuottajien prosessitehokkuutta. Haastattelussa moni koki tällaisen moninkertaisen työn vähentämisen erittäin positiiviseksi.

AV-alan jakeluvaihtoehdot lisääntyvät. Liiketoimintaprosesseista tulee koko ajan data-intensiivisempiä ja automatisoidumpia. Tämä edellyttää standardointia ja pysyviä tunnisteita. On tärkeää, että myös suomalainen AV-ala on mukana tässä kehityksessä. Päätöksenteossa käytettävä tieto muodostaa kilpailuetua. Tuon tiedon synnyttäminen vaatii dataa ja osaamista, joita tähän asti AV-alalla ei ole käytetty laajemmassa mittakaavassa.



Kuva 9: Keskitetty datan syöttö tuottajille



Kuva 10: Tiedonsiirron keskittäminen yhteen solmukohtaan

Haastatellut henkilöt

Haastateltu	Asema	Organisaatio	Päivämäärä
Anna Lujanen	Toiminnanjohtaja	Screenforce Finland	26.3.
Kim Viljanen	Digitaalisen tuotekehityksen, sisällönhallinnan ja linkitetyn metatiedon asiantuntija	YLE	29.3.
Heidi Keinonen	Perustaja	Medialogi Oy	31.3.
Eerika Vermilä	Perustaja	Medialogi Oy	31.3.
Mika Toivainen	Taluspäällikkö	Endemol Shine Finland	31.3.
Magnus Pettersson	Director Of Development, CX and Service Portfolio	Elisa	1.4.
Turo Pekari	Senior advisor, Data transformation	Teosto	1.4.
Ville Mäkinen	Päällikkö, oikeudenomistajat	Teosto	1.4.
Ilkka Tuomela	Palvelupäällikkö	YLE	12.4.
Elina Selkälä	Arkistopäällikkö	YLE	12.4.
Pia Virtanen	Product Owner	YLE	12.4.
Pasi Ekman	Päällikkö, Kehitys ja tuki	YLE	12.4.
Meira Noronen	Operatiivinen johtaja (COO) / päällikkö, non-scripted	Endemol Shine Finland	12.4.
Suvi Salokoski	Päällikkö, non-scripted	Warner Bros Finland	13.4.
Petri Haverinen	ICT ja liiketoiminta applikaatioiden päällikkö	Telia	15.4.
Jan Tanhuanpää	Suunnittelija	Kopiosto	26.4.
Vallius Rainer	Tilitys- ja tutkimuspäällikkö	Kopiosto	26.4.
Lena Brun	Toimitusjohtaja	Finnpanel	6.5.
Antti Penttala	Tuotantojohtaja, Data ja raportointi	Finnpanel	6.5.
Régis Flad	Toimitusjohtaja	ISAN	6.5.
Tero Koistinen	Toimitusjohtaja	Filmikamari	6.5.
Matti Lampinen	Johtaja, Tekniset projektit ja jakelu	Sanoma	20.5.
Marcus Wiklund	Viestintä- ja yhteiskuntasuhdejohtaja	Sanoma	20.5.
Miikka O'Connor	Co-Founder, COO	Pongfinity	22.6.

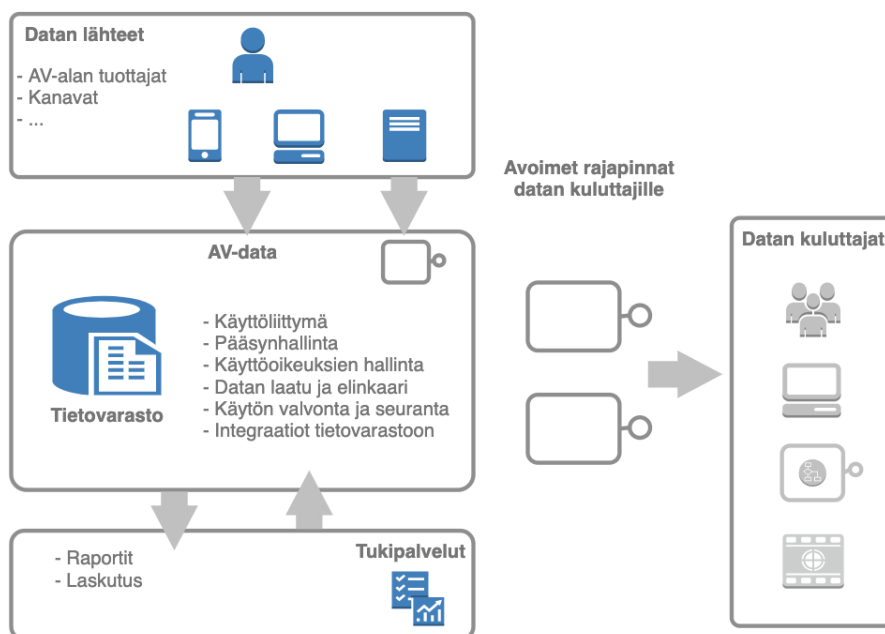
Portaalin tekninen kuvaus

1 Yleiskuva

1.1 Yleiskuva teknisestä ratkaisusta

Audiovisuaalisen alan eri toimijoiden yhteiseen käyttöön soveltuva ja valikoitunut tieto kerätään keskitettyyn tietovarastoon, josta se julkaistaan jäsenten käyttöön soveltuvien käyttöoikeuksien mahdollistamassa laajuudessa. Ratkaisussa on huomioitu tarve käyttöoikeuksien rajaamiseen eri tasoisesti eri toimijoiden kesken, tarkempi kenttätasoinen suunnittelu näkyvyydestä tehdään myöhemmin.

Datan syöttöön ja rikastamiseen tarjotaan sekä käyttöliittymä että tekniset rajapinnat. Datan osia on mahdollista tarjota myös verkoston ulkopuolisille tahoille rajapintojen kautta näin sovittaessa.



Kuva 1. AV-alan datan ekosysteemi

1.2 Ylätason palvelut

AV-dataverkoston ydin koostuu seuraavista palveluista:

1. **Tietovarasto** – varsinainen tiedon tallennuspaikka, joka toteutetaan joustavalla tietomallilla, jotta se palvelee mahdollisimman hyvin verkoston jäsenten tarpeita ja on laajennettavissa tulevaisuuden tarpeisiin.
2. **Käyttöliittymä** – tietoa voidaan syöttää, hakea ja täydentää eri päätelaitteilla. Käytön helppouteen ja tietojen syötön joustavuuteen kiinnitetään erityistä huomiota.
3. **Rajapinnat dataan** – verkoston jäsenet voivat hyödyntää teknisiä rajapintoja sekä tietojen haakuun omiin järjestelmiinsä että tietojen tallennukseen yhteiseen käyttöön omista

järjestelmistään. Dataan voidaan tarjota rajapinnat myös verkoston ulkopuolisille jäsenille esimerkiksi laskutettavina palveluina.

4. **Pääsynhallinta ja käyttöoikeudet** – tiedon näkyvyyttä ja muokkausoikeuksia rajataan rooli-pohjaisiin käyttöoikeuksiin.
5. **Datan laadun valvontaan** tarjotaan mekanismit, jolla alkuperäinen lähde ja muutoshistoria on helposti nähtävissä. Samasta tietosisällöstä voi olla rinnakkaisia versioita lähteestä riippuen.
6. **Tukipalveluita** kehitetään tarpeen mukaan esimerkiksi raportointiin ja datan käytön laskutukseen.

2 Järjestelmän arkkitehtuuri

2.1 Tekniset valinnat ja periaatteet

Datavarasto toteutetaan keskitettynä pilvipohjaisena ratkaisuna. Tässä materiaalissa kuvattu ratkaisu on toimittajariippumaton – kaikilla johtavilla pilvipalveluiden tarjoajilla (Amazon Web Services, Microsoft Azure ja Google Cloud Platform) löytyy omat tuotteet tekniselle toteutukselle. Kappaleessa 5 oleva kustannusarvio on laskettu yhdellä vaihtoehdolla, mutta koska kyse on pilviarkkitehtuurin peruskomponenteista, ei kustannuksissa yleensä ole merkittäviä eroja toimijoiden välillä.

Vastakohtana keskitetylle ratkaisulle voisi olla dataverkoston jäsenten kesken hajautettu datavarasto, jossa datan eri osat kerättäisiin ja tallennettaisiin dataverkostoon kuuluvien kumppanien järjestelmissä. Tämä aiheuttaa kuitenkin monimutkaisemman ja vaikeasti hallittavan kokonaisuuden, jonka ylläpito ja laadunhallinta ovat merkittävästi raskaampi.

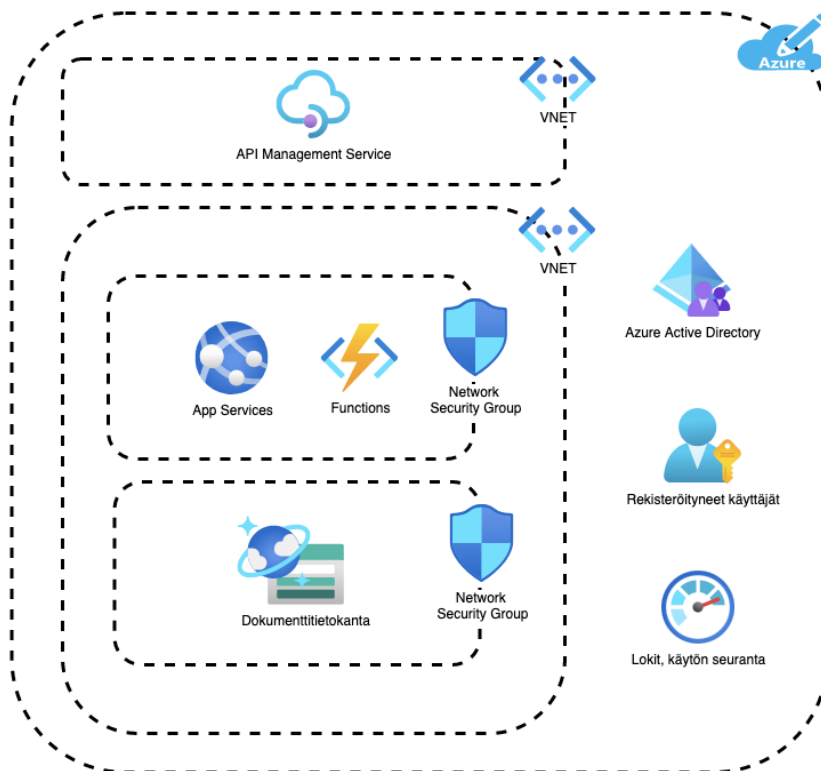
Periaatteena on, että kertaalleen syötetty tieto on heti muiden verkoston jäsenten käytettävissä (käyttöoikeuksien rajoissa) ja tietoa voidaan täydentää vaiheittain.

2.2 Tekninen ratkaisukuvaus

Tietovarasto perustuu dokumenttietokantaan, jotta se on joustava ja mahdollisimman helposti laajennettavissa tietomallin osalta myöhemmin. Tietovarastoon voidaan tallentaa myös ääni-, valokuva- ja videotiedostoja kustannustehokkaasti.

Kaikki tiedon syöttöön ja hakuun tarjottavat palvelut perustuvat REST-pohjaisiin mikropalveluihin, jotka tarjotaan jonkin API-hallintatyökalun kautta (esim. Azure-ympäristössä API Manager). API:en hallintatyökalun avulla voidaan keskittää tietoturva, skaalautuvuus, rajapintojen versiointi, käyttäjien dokumentaatio yms. Erityisesti tiedon näkymiseen vain oikeille käyttäjille kiinnitetään erityistä huomiota.

Käyttöliittymästä on toteutettu demoversio React-viitekehysellä, ja sitä voidaan käyttää ratkaisun esittelyssä. Demo-versio ei tallenna tietoa taustapalveluun. Demo on asennettuna osoitteeseen <https://data-alpha.vercel.app>.



Kuva 2: Tietovarasto ja palvelut – tekninen rakenne esimerkiksi Azuren komponenteilla toteutettuna.

2.3 Tietoturvaperiaatteet

1. Järjestelmän käyttö edellyttää aina kirjautumista, joka perustuu sähköpostitunnukseen ja siihen lähettyyn kertakäyttöiseen salasanaan.
2. Järjestelmän käyttö suoraan rajapintojen kautta edellyttää sitä varten myönnettyä API-avainta.
3. Järjestelmä noudattaa roolipohjaista käyttöoikeuksien hallintaa.
4. Roolipohjaista käyttöoikeuksien hallintaa täydennetään sisältöpohjaisella (metadataan perustuvalla) käyttöoikeuksien hallinnalla.
5. Tietojen haut ja tallennukset tapahtuvat aina palvelukerroksen kautta jossa tehdään tarvittava lokitus.
6. Kaikki tietoliikenne ja tallennettu tieto on salattua.
7. Toistuvat epäonnistuneet kirjautumisyritykset passivoivat käytetyn tunnuksen, kaikki tekniset tunnistetiedot kirjautumisyrityksestä kirjoitetaan lokille (käyttäjätunnus, käyttäjän ip-osoite, käytetty pääte-laite jne).
8. Ennen käyttöönottoa järjestelmälle tehdään tietoturva-auditointi ja penetraatiotestaus.

3 Toiminnalliset ja ei-toiminnalliset vaatimukset

Alla olevassa taulukossa on listattu järjestämän vaatimukset otsikkotasolla.

1.1. Perustoiminnallisuus	
1	Ratkaisulla voidaan kerätä AV-alan dataa yhteiskäyttöiseen tietovarastoon
2	Datan hallinta ja näkyvyys on helposti rajattavissa käyttäjäkohtaisesti
3	Dataa voidaan syöttää ja muokata vaiheittain
4	Järjestelmä tukee käyttäjää tiedon syöttämisessä niin, että sama tieto syötetään vain kerran
5	Järjestelmä tukee käyttäjää tiedon syöttämisessä niin, että se tarjoaa ennakoivaa listaa mahdollisista vaihtoehdoista
1.2. Käyttöliittymä ja rajapinnat	
1	Järjestelmää voidaan käyttää yleisimmillä selaimilla ja eritasoisilla päätelaitteilla (tietokone, tabletti, älypuhelin)
2	Järjestelmärajapinnat on toteutettu OpenAPI 3 -määritysten mukaisesti.
3	Tiedon syöttämisen sujuvuuteen on kiinnitetty erityistä huomiota
2.1. Tietoturva ja datan tallennus	
1	Data säilytetään EU:n alueella
2	Data on salattua sekä tallennuksen että siirron aikana
3	Järjestelmä tukee roolipohjaista käyttöoikeuksien hallintaa
4	Järjestelmä tukee sisältöpohjaista (metadatatapohjaista) käyttöoikeuksien hallintaa
5	Järjestelmässä on automaattinen tietojen varmuuskopiointi omalle medialle tai palvelimelle, pois käytetystä palveluympäristöstä.
2.2. Skaalautuvuus ja suorituskyky	
1	Järjestämään tallennettava data vuositasolla on n. 200 Gb, ks. lopussa kappale 5.2.
2	Järjestelmään rekisteröityneitä käyttäjiä on noin 300

4 Data

4.1 Datan sijainti ja elinkaari

Data varastoidaan EU-alueella olevassa datakeskuksessa. Vanhaa dataa ei poisteta, mutta sitä voidaan arkistoida automaattisesti halvemman varastoinnin tasolle kustannusten säästämiseksi. Tällöin datan säilyttäminen on halvempaa, mutta vastaavasti kaikki siihen kohdistuvat luku- ja kirjoitusoperaatiot maksavat enemmän.

Data voidaan luonnollisesti myös varmuuskopioida automaattisesti ulkoiselle medialle ”omaan talteen”, ulos pilvipohjaisesta ratkaisusta.

4.2 Datan suojaus ja käytön seuranta

Kaikki tallennettu data on salattua sekä varastoituna (”at-rest”) että luku- ja kirjoitusoperaatoiden aikana (”in-transit”).

Tehdyt tallennukset ja haut voidaan kirjoittaa lokille (kuka käyttäjä tallensi tai haki jne.) – lokille kirjoitettavat toiminnot ja tietoalkiot yksilöidään tarkemman järjestelmän suunnittelun yhteydessä.

4.3 Käyttöoikeudet ja roolit

Datan näkyvyys suojataan sekä rooliin pohjautuvilla että sisältöön (metadataan) pohjautuvilla käyttöoikeuksilla. Roolipohjaiset käyttöoikeudet (RBAC, Role Based Access Control) on yleinen tapa ja se toimii käyttöoikeuksien perustana, mutta sitä täydennetään ja tarkennetaan sisältöön pohjautuvalla pääsynhallinnalla.

Rooleja, joiden puitteissa oikeuksia myönnetään, voi olla useita, ja niitä hallitaan käyttäjäryhmien kautta, esimerkiksi:

- Ryhmä ”Datan omistaja yrityksessä A”: kaikki tähän käyttäjäryhmään liitetyt käyttäjät saavat kaikki oikeudet yrityksen A dataan.
- Ryhmä: ”Kirjoitusoikeus yrityksessä A”: tähän käyttäjäryhmään liitetyillä käyttäjillä on kirjoitus ja lukuoikeus yrityksen A dataan, mutta ei esimerkiksi poisto-oikeutta
- Ryhmä: ”Lukuoikeus yrityksessä A”: tähän käyttäjäryhmään liitetyt käyttäjät voivat vain lukea lukea yrityksen A dataa.

Mikäli käyttäjä ei ole missään yllä olevissa ryhmissä, pääsyä yrityksen A dataan ei ole. Sama käyttäjä voidaan liittää useaan eri ryhmään jolla on eritasoiset oikeudet. Esim. hänet voidaan liittää ryhmään ”Datan omistaja yrityksessä A” ja ”Lukuoikeus yrityksessä B”. Datan näkyvyys ei tarvitse olla rajoittunut yritysta-

soisiin tietoihin, vaan se voi olla myös kenttäkohtaista. Voidaan esimerkiksi määritellä, että nähdäkseen koko henkilötunnuksen tai sen loppuosan, pitää kuulua ryhmään "Oikeus käsitellä henkilötietoja".

Sisältöön pohjautuvassa pääsynhallinnassa roolin lisäksi pääsy dataan on rajattu sen sisällön perusteella. Esimerkiksi jos käyttäjä on liitetty ryhmään, jolla on lukuoikeus yrityksessä C, hän näkee lähtökohtaisesti yrityksen C julkaisemat teokset ja niiden tiedot. Erikseen voidaan kuitenkin rajata yksittäisen teoksen osalta, että jos masternauhan suunniteltu toimituspäivä on tulevaisuudessa, mitään tietoa nimenomaan tästä teoksesta ei tälle käyttäjälle näytetä.

5 Kustannukset

5.1 Kehityskustannukset

Alla olevaan taulukkoon on eritelty suuntaa-antava arvio järjestelmän vaatimista kehityspanoksesta henkilötyöpäivinä (htp).

Työvaihe	Arvio
<p>Järjestelmän vaatimusmäärittelyjen tarkentaminen yhdessä kaikkien sidosryhmien kanssa, reilun datatalouden sääntökirjan laatiminen</p> <ul style="list-style-type: none"> • haastattelut, 2 h / sidosryhmä, 7 sidosryhmää = 2 htp • vaatimusten tarkennukset 2 htp • reilun datatalouden sääntökirja 5 htp 	9 htp
<p>Käyttöliittymädemon täydentäminen valmiiksi sovellukseksi</p> <ul style="list-style-type: none"> • Kirjautuminen 1 htp • Uusien kenttien ja toiminnallisuuden lisääminen vaatimusmäärittelyn perusteella 3 htp • Käytettävyyden viimeistely 2 htp • Kenttien opastustekstit ja syötetyn sisällön validointi 1 htp • Hakutoiminnallisuus ja hakunäyttö jo tallennettuun dataan 2 htp • Tallennus keskitettyyn tietovarastoon 2 htp 	11 htp

<p>Käyttäjien- ja pääsynhallinta</p> <ul style="list-style-type: none"> • Käyttäjätunnusten hallinta 1 htp • Kertakäyttösalasana -toiminnallisuus 1 htp • Käyttöoikeuksien hallinta ja konfigurointi vaatimusten mukaan 3 htp 	5 htp
<p>Tietovarasto</p> <ul style="list-style-type: none"> • Dokumenttitietokanta 3 htp • API-rajapinta käyttöliittymälle 2 htp • API-rajapinnan tietoturvan toteutus ja lokitukset 2 htp • Automaattinen varmuuskopiointi 1 htp 	8 htp
<p>API-rajapinta kumppaneille</p> <ul style="list-style-type: none"> • Lisäykset edellä olevaan käyttöliittymän käyttämään rajapintaan 2 htp • Rajapinnan OpenAPI 3 dokumentaatio ja sen julkaisu kumppaneille 3 htp 	5 htp
<p>Ylläpidon ja pääkäyttäjän ohjeet sekä koulutus 2 htp</p>	2 htp
<p>Yhteensä</p>	40 htp

5.2 Käyttökustannukset

Tämän tyyppisessä ratkaisussa edullisin vaihtoehto on valita pilvipalvelun toimittajalta käyttöpohjainen laskutusmalli, jossa kaikki kustannukset perustuvat käyttömäärään. Toisaalta juuri siitä johtuen tarkka kustannusten ennakointi on vaikeaa, koska mitään käyttötilastoa ei laskennan pohjaksi ole. Alla on esitetty suuntaa antava arvio Microsoft Azure -ympäristössä. Azure-ympäristössä dokumenttitietokannan laskennan pohjana käytetään suuretta Request Unit, RU, jonka määrittelystä on tarkemmin esim. [täällä](#).

Jos:

- teoksia tallennetaan noin 200 000 kpl kymmenen vuoden aikana,
- niiden keskimääräinen koko dokumenttitietokannassa olisi 10 kt
- erikseen teokseen liittyvien liitteiden (kuvat, videot) koko olisi 10 Mt per teos
- teoksia haettaisiin 1000 kertaa vuorokaudessa erilaisilla hakutermeillä
(10 RU/haku * 1000 haku/vrk * 31 vrk = 310 000 RU)
- teoksia ladataan sovellukseen yhtä monta kertaa
(1 RU/kt * 10 kt * 1000 latausta/vrk * 31 vrk = 310 000 RU)
- teoksia päivitetäisiin 50 kertaa vuorokaudessa
(10 RU/kt * 10 kt * 50 päivitystä / vrk * 31 vrk = 155 000 RU)

Kustannukset muodostuisivat seuraavasti:

1. Varsinainen sovellus, joka tarjoaa demon kaltaisen sovelluksen:
Azure App Service, B1 taso, kustannus **11,08 €/kk.**
2. Dokumenttitietokanta, rajapinnat ja dokumenttien varastointi:
 - a. Cosmos DB, Serverless -versio, rajapinta
775 000 RU / 1 000 000 * 0.239 €/1M RU = **0,19 €/kk**
 - b. Cosmos DB, Serverles -versio, tallennus
200 000 dokumenttia * 10 kt * 0,211 € / 1 Gt = **0,42 €/kk**
3. Liitetietokanta
 - a. Azure Blob Storage
200 000 dokumenttia * 10 Mt * 0,0186 € / 1 Gt = **37,2 €/kk**
(Huom. 10 vuoden dokumenttimäärällä).

Eli järjestelmän käyttökustannukset kuukaudessa olisivat noin **50 €** kymmenen vuoden dokumenttimäärällä arvioituna. Epävarmuustekijöitä (dokumenttien määrä ja koko, hinnoittelun muuttuminen jne.) on kuitenkin paljon, joten arvio on syytä vähintään kaksinkertaistaa.

Muita mahdollisia käyttökustannuksia voivat aiheuttaa lisäksi

- sähköpostipalvelimen käyttö (kertakäyttösalasanojen lähetys)
- varmuuskopioiden tallennus ja datan siirto "ulos" pilviympäristöstä