

KAVIn Ritva-tietokanta tutkimusdatan keräämisen, käytön ja hallinnan välineenä

Eeva Savolainen

Digitaalista tutkimuskäyttöön sopivaa dataa on nykyisin tarjolla yhä enemmän, minkä vuoksi datan hallintaan tarvitaan tehokkaita työkaluja. Artikkelissa tarkastellaan, kuinka tutkijoille suunnattu digitaalinen palvelu voi tukea tutkimusdatan hallinnan eri vaiheita.

Digitaalinen aikakausi on avannut uusia mahdollisuuksia muistiorganisaatioille, joiden koekelmat ja metadata ovat yhä useammin digitaalisessa muodossa. Tutkijoille kehitys on tuonut pääsyn yhä laajempaan määrään tutkimuskäyttöön soveltuvaa dataa. Datan runsaus aiheuttaa myös ongelmia: unelma kaiken saatavilla olevan datan löytämisestä saattaa muuttua painajaiseksi hukkumisesta tiedon mereen (ks. Veyrat-Masson 2012). Datan hallinnasta onkin tullut kestoaihe sekä tutkimus- että kulttuuriperintöorganisaatioille.

Artikkelissani tarkastelen Kansallisen audiovisuaalisen instituutin (KAVI) radio- ja tv-arkiston Ritva-tietokantaa tutkimusdatan keräämisen, käytön ja hallinnan työkaluna. Analysoimalla Ritvan heikkouksia ja vahvuuksia pyrin osoittamaan, miten digitaalinen palvelu voi tukea tutkimusdatan hallinnan eri vaiheita, ja missä asioissa on vielä kehittämistä. Tutkimusdatalla viittaa tässä yhteydessä radio- ja tv-sisältöön sekä siihen liittyvään metadataan.

Radio- ja tv-arkistointi Suomessa

Radio- ja tv-aineistoa on arkistoitu Suomessa 1950-luvun lopulta lähtien. Yleisradion filmiarkisto aloitti toimintansa vuonna 1958 ja radioarkisto 1961 (Selkälä 2016). Myös kaupalliset kanavat ovat ylläpitäneet omia arkistojaan. Kansallista radio- ja tv-arkistoa jouduttiin kuitenkin odottamaan aina vuoteen 2008, jolloin uuden kulttuu-

riaineistolain (L 28.12.2007/1433) myötä Suomen elokuva-arkiston yhteyteen perustettiin radio- ja tv-arkistointiin keskittyvä yksikkö. Uuden organisaation nimeksi tuli Kansallinen audiovisuaalinen arkisto (vuodesta 2014 Kansallinen audiovisuaalinen instituutti).

Radio- ja tv-arkistolla on kaksi päätehtävää. Ensiksikin se tallentaa ohjelmavirtaa tutkimuskäyttöön (Ritva-tietokanta), ja toiseksi se tallentaa kaupallisten kanavien alkuperäislaatuisia ohjelmatalenteita kulttuuriaineistolaissa määritellyistä ohjelmatyypeistä (esim. dokumentit ja draama). Yleisradio hoitaa edelleen omilla kanavillaan lähetettyjen alkuperäislaatuisten ohjelmatalenteiden tallettamisen ja säilytyksen.

Ritva-tietokannan sisältö, rakenne ja metadata

Ohjelmavirran tallentaminen Ritva-tietokantaan alkoi vuoden 2009 alussa. Valtakunnalliset kotimaiset radio- ja tv-kanavat tallennetaan kokonaisuudessaan, paikalliskanavista viikon näytteet vuosittain. Tietokannan historiallinen ulottuvuus laajenee huomattavasti vuonna 2016, kun siihen lisätään Yleisradion digitoidut tv-ohjelmat 1960-luvulta vuoteen 2008. Tietokanta on käytettävissä KAVIn lisäksi nk. katselu- ja kuuntelupisteillä, jotka on suurimmaksi osaksi sijoitettu vapaakappalekirjastojen yhteyteen.

Ritva-tietokanta on relaatiotietokanta, joka koostuu toisiinsa linkitetyistä tietueista. Keskei-

Yle Uutiset

Ohjelma

Ohjelman nimi	Yle Uutiset	Lähde	yle metadata
Ohjelman nimi	Yle Uutiset		finnpaneli
Ohjelman nimi	Yle Nyheter		yle metadata
Genre (Venetsia)	Uutiset		venetsia
Genre (EPG)	Uutis- ja ajankohtaisohjelmat		epg
Genre (Finnpanel)	Uutiset		finnpaneli
Pysyvä tunniste	http://rtva.kavi.fi/program/details/program/8404527		

Tekijät

- Hartikainen, Aleksj, Apulaisohjaaja
- Junell, Ulla, Apulaisohjaaja
- Simmola, Marjut, Ohjaaja
- Iskakaas, Lasse, Graafikko
- Bergstrom, Anne, Meteorologi
- Kentia, Ilmari, Uutispäällikkö
- Frieland, Derrick, Kuvatoimittaja
- Franji, Tommy, Lukija (kertoja/speak)

Pillota

Lähetys

- Yle TV1, 12.8.2012 klo 18.00-18.10 [Lähetysten tiedot](#) [Näytä kanava ohjelmaoppaassa \(EPG\)](#) [Näytä aikajana](#)
- Toista lähetys [Pillota pikkukuvat](#)
- Näytä: 6 kuvaa 1 min [Kaikki kuvat](#)

Kuva 1.

simmat tietuetyypit ovat ohjelma-, jakso-, sarja-, lähetys-, kanava-, henkilö- ja yhtiötietue. Kuvassa 1 on ohjelmatietue, jossa on linkit henkilö-, lähetys- ja kanavatietueisiin.

Tietokannan metadata koostuu eri lähteistä importoiduista tiedoista ja arkiston tuottamasta metatiedosta. Keskeisiä importoidun metadatan lähteitä ovat kotimaisten radio- ja tv-kanavien ohjelmatiedotusjärjestelmä Venetsia, kanavien ohjelmaoppaat (EPG), kotimaista radiokuuntelua ja television katselua mittaava Finnpanel Oy, kotimaisten radiokanavien yhteinen tietojärjestelmä Radiot.fi sekä Yleisradio. Kuvassa 1 näkyy, että kyseiseen tietueeseen on saatu tietoja Yleisradiolta, Finnpanelilta, Venetsiasta ja EPG:stä. Tämän lisäksi arkisto tuottaa tietokantaan metadataa, jota ei saada tiedonsiirtoina. Tällaisia tietoja ovat esimerkiksi kotimaisten sarjojen sisällönkuvaukset ja kanavakuvaukset, jotka sisältävät tietoa kanavan historiasta ja ohjelmistosta.

Ritva tutkimusdatan keräämisen välineenä

Tutkijan kannalta jatkuva ohjelmavirran tallentaminen on merkittävä parannus aineistojen ke-

räämiseen. Radio- ja tv-ohjelmatalenteet olivat aiemmin ohjelmaa lähettävien yhtiöiden, tuotantoyhtiöiden ja yksityisten henkilöiden hallussa, mikä edellytti tutkijalta sala-poliisintaitoja aineistojen jäljittämiseksi. Lisäksi talenteita oli usein mahdollista käyttää vain kyseisen yhtiön tiloissa. (Keinonen 2012, 54–55.) Ritva-tietokannan kautta tutkija pääsee käsiksi kaikkien keskeisten kotimaisten kanavien ohjelmavirtaan vuoden 2009 alusta lähtien, ja pian myös Yleisradion di-

gitoituihin vanhempiin tv-ohjelmiin. Paikalliskanavien aineistoihin pääsy on kuitenkin edelleen hankalaa: niiltä on saatavilla vuosittain vain viikon näytteet. Ritvassa ei myöskään ole toistaiseksi kaupallisten kanavien historiallista aineistoa. Joka tapauksessa pääsy radio- ja tv-aineistoon on nykyisin huomattavasti helpompaa kuin aiemmin.

Ohjelmavirtaa on tallennettu vuoden 2009 alusta lähtien systemaattisesti ja kattavasti. Aiemmin yhtiöt päättivät itse, mikä oli säilyttämisen arvoista (vrt. esim. Noordegraaf 2010, 7). Monet tutkijoiden kannalta kiinnostavat ohjelmat jäivät arkistoinnin ulkopuolelle. Nykyisin vastaavanlaisia aukkoja ei pääse enää syntymään. Tutkijat, jotka etsivät tulevaisuudessa radio- ja tv-aineistoa tutkimuksensa kohteeksi, eivät siten ole enää tutkimusaiheita valitessaan sidottuja yhtiöiden arkistointipolitiikkaan.

Tutkijoiden työtä helpottaa myös se, että arkisto huolehtii sekä aineiston tallennuksesta että pitkäaikais säilytyksestä. Aineiston käyttö on ilmaista ja siihen pääsee käsiksi eri puolilla Suomea vapaakappalekirjastoissa.

Ohjelmaopas EPG / Aikajana	
Lisää kanava: <input type="text"/>	Valitse päivä: 2.1.2016
MTV3	Nelonen
0:22-0:26 Eurojackpot ja Jokeri	1:36-1:40 Tulossa: Mustat lesket
0:28-1:21 Mustamaalattu (12)	1:40-2:28 Zero Hour (12)
1:24-5:59 Voittostudio	2:34-2:35 Tulossa: Ihon alla

Kuva 2.

Ohjelmaopas EPG / Aikajana	
Lisää kanava: <input type="text"/>	Valitse päivä: 2.1.2016
MTV3	Nelonen
12:30:00	13:30:00
Ohjelmat	Ohjelmat
11:30-13:24 Valtamerten pinnan alla (5)	13:24-14:54 Big Top Scooby-Do! (7)
12:30:00	13:30:00
14:30:00	15:30:00

Kuva 3.

Ritva tutkimusdatan käytön näkökulmasta

Haku

Tutkimusdatan käytön näkökulmasta Ritva-tietokanta tarjoaa tutkijalle monia etuja. Tietokanta edistää välinerajojen tutkimuksellista ylittämistä mahdollistamalla samaan aikaan sekä radio- että tv-aineistoon kohdistuvat haut (Keinonen 2012, 56). Monipuoliset hakutyökalut mahdollistavat relevantin datan löytymisen laajan digitaalisen aineiston joukosta. Hakuja voi rajata kuvailevan metadatan mukaan (esim. ohjelman nimi, kuvaus, asiasana, genre), kanavan mukaan, ohjelmatyyppin mukaan (ohjelma, sarja, jakso), mediatyyppin mukaan (radio tai televisio), henkilön tai yhtiön mukaan. Hakutuloksia on mahdollista ladata eri formateissa (esim. csv ja xml) jatkokäsittelyä varten.

Selaus

Hakutyökalujen lisäksi aineistoa on mahdollista myös selata. Kanavia, henkilöitä ja yhtiöitä voi selata nimen mukaan, teksti-tv-sivuja kanavittain tai päivän mukaan. Ohjelmaopassivu EPG (ku-

va 2) näyttää valittujen kanavien ohjelmatarjonnan tietyltä päivästä. Ohjelmien nimet toimivat linkkeinä ohjelmatietueisiin ja tallenteisiin.

Aikajananäkymällä (kuva 3) voi verrata eri kanavien ohjelmavirtaa rinnakkain visuaalisena näkymänä.

Audio/videotallenne-näkymä (kuva 4) esittää yhteenvedon siitä, mitä viikkoja kanavilta on tallennettu vuosittain.

Haku- ja selaustoiminnoissa on myös haasteita. Tehokkaiden tiedonhaku- ja selaustoiminnoissa on edellyttävä käyttäjältä melko hyvää tietokannan rakenteen, metadatan ja sisällön tuntemusta. Toistaiseksi hakuja ja hakutuloksen latauksia ei voi tehdä muualla kuin katselu- ja kuuntelupisteillä, mutta tämä on muuttumassa. Metadatabasit on tarkoitettu julkaista internetissä vuoden 2016 aikana, minkä jälkeen hakuja voi tehdä verkossa vapaasti ja kirjautuneet käyttäjät voivat myös ladata hakutuloksia.

Katselu ja kuuntelu

Ritva-tietokantaa varten on kehitetty selaimen upotettu, räätälöity VLC-soitin. Ohjelmavirrassa voi navigoida esimerkiksi soittimeen integroitujen pikkukuvia klikkaamalla tai ohjelmaoppaan, kalenterin tai kanavavalikon avulla.

Ongelmallista on se, että katselu ja kuuntelu on tekijänoikeusyistä rajattu vain tiettyihin organisaatioihin. Digitaalinen kopiointi ja välittäminen on samasta syystä estetty katselu- ja kuuntelupisteillä, joten tutkija ei voi tallentaa haluamaansa ohjelmavirtaa omalle koneelleen. Internetjulkaisukaan ei auta tässä, sillä vain metadata voidaan

Audio/Videotallenteet

Valitse vuosi: 2016 Kanavat: TV Näytä

Yksi täysi tallennusviikko = 168 tuntia. Taulukon luvut esittävät viikon a

	1	2	3	4	5	6	7	8	9	10	11
Yle TV1	168	168	168	168	168	169	168	168	168	168	168
Yle TV2	168	168	168	168	168	169	168	168	168	168	168
Yle Fem	168	168	168	168	168	169	168	168	168	168	168
Yle Teema	168	168	168	168	168	169	168	168	168	168	168
MTV3	168	168	168	168	168	169	168	168	168	168	168

Kuva 4.

julkaista verkossa tekijänoikeussyistä. Ajoittaiset tekniset ongelmat ja tallennuksen häiriöt saattavat myös ajoittain häiritä aineiston käyttöä.

Ritva tutkimusdatan hallinnan välineenä

Ritva tarjoaa työvälineitä myös tutkimusdatan hallintaan. Työpöytätyökalulla käyttäjä voi tallentaa tekemiään hakuja, kiinnostavia ohjelmapätkiä ja näihin liittyviä muistiinpanoja. Ohjelmatallenteita lukuun ottamatta näihin pääsee jatkossa käsiksi myös internetissä, mutta toistaiseksi tutkijat joutuvat tekemään muistiinpanonsa omalla koneellaan.

Runsa metadatan määrä voi tukea tutkimusdatan dokumentointia. Arkiston tuottama metadata kuten sarjakohtaiset kuvaukset ja kanavakuvaukset voivat tarjota apua tiedon kontekstualisoinnissa. Hakutuloksia ja metadataa voi ladata eri formaateissa jatkokäyttöä varten. Jokaisella tietueella on pysyvä tunniste, joka auttaa aineistoon viittaamisessa.

Vaikka metadataa on paljon, sitä on tarjolla paikoin epätasaisesti. Jotkut tietueet sisältävät vain perustiedot: joissakin ohjelmatietueissa on esimerkiksi vain nimi ja lähetyisaika. Toisissa tietueissa kuten kotimaisissa sarjatieueissa on puolestaan runsaasti metadataa käytettävissä: ohjelman eri nimiversioita, sisältökuvauksia, tietoa tuotantovuodesta, asiasanoja sekä linkityksiä henkilöihin ja yhtiöihin.

Metadatassa, esimerkiksi henkilö- ja yhtiölinki-

tyksissä, on paikoin aukkoja. Yleltä saadaan henkilö- ja yhtiölinkityksiä osaan ohjelmista, ja kotimaisiin sarjoihin lisätään linkkejä, mutta resurssit eivät riitä linkittämään näitä tietoja esimerkiksi yksittäisiin sarjan jaksoihin. Suuri osa tiedonsiirtoina saaduista ohjelmien sisältökuvauksista on alun perin luotu markkinointitarkoituksiin, joten niiden soveltuvuus tutkimuksen tiedonlähteeksi voi olla paikoin kyseenalainen. Lisäksi metadata saattaa sisältää virheitä, sillä arkistossa ei ehditä käymään lävitse kaikkea tiedonsiirtoina saatua metadataa.

si metadata saattaa sisältää virheitä, sillä arkistossa ei ehditä käymään lävitse kaikkea tiedonsiirtoina saatua metadataa.


Yhteenveto

Kaikkiaan Ritva-tietokanta tarjoaa apukeinoja tutkimusdatan hallintaan, vaikka kehitettävää toki riittää. Tietokanta tarjoaa laajan ja kattavan pääsyn kotimaiseen radio- ja tv-aineistoon. Arkisto vastaa aineiston tallennuksesta ja pitkäaikais säilytyksestä säästäten siten tutkijan työtä. Tietokannan käyttö on maksutonta ja siinä on useita haku- ja selailuominaisuuksia sekä räätälöity työkalu aineiston katseluun ja kuunteluun. Raakametadataa sekä arkiston tuottamaa metadataa on tarjolla runsaasti. Pysyvät tunnisteet helpottavat aineistoon viittaamista.

Vaikka tietokanta sisältää paljon aineistoa, siinä on edelleen sisällöllisiä puutteita. Paikalliskannalta on käytettävissä vain viikon näytteitä. Historiallisessa mielessä Ritva-tietokanta on toistaiseksi melko kapea-alainen. Ohjelmavirtaa on tallennettu vuoden 2009 alusta ja historiallista aineistoa edustavat toistaiseksi lähinnä Ylen digitoitut ohjelmat. Tulevaisuudessa Ritva-tietokantaan on kuitenkin mahdollista liittää myös muita historiallisia kokoelmia.

Keskeisin käyttöä hankaloittava tekijä on katselun ja kuuntelun rajautuminen vain tiettyihin organisaatioihin. Katselun ja kuuntelun vapauttaminen internetissä edes tutkimuskäyttöön vaatisi kuitenkin nykyiseen lainsäädäntöön muutoksia, joita toistaiseksi ei ole näköpiirissä. Metadataa saadaan eri ohjelmista ja eri kanavilta epäta-

saisesti, ja resurssit eivät riitä paikkaamaan kuin osan metadatan aukoista. Lisäksi metadata saattaa sisältää virheitä, koska kaikkea metadataa ei ehditä tarkistaa.

Toistaiseksi hakujen tekeminen sekä hakutulosten ja muistiinpanojen käsittely vaatii asiakasta hakeutumaan katselu- ja kuuntelupisteelle, mutta internetjulkaisun jälkeen nämä haasteet jäävät historiaan. Arkisto pyrkii jatkuvasti etsimään ratkaisuja myös muihin tietokannan rakenteeseen ja käyttöön liittyviin haasteisiin, jotta tietokanta olisi jatkossa yhä parempi työkalu myös tutkimusdatan keräämiseen, käyttöön ja hallintaan. 

Lähteet:

Keinonen, Heidi (2012). KAVA:n radio- ja tv-arkisto: ohjelmavirtaa ja talletuksia tutkimuskäyttöön. *Lähi-kuva* 1/2012, 53-57.

L 28.12.2007/1433. Laki kulttuuriaineistojen tallettamisesta ja säilyttämisestä. Valtion säädöstietopankki Finlex, Ajantasainen lainsäädäntö. <http://www.finlex.fi> (12.7.2016)

Noordegraaf, Julia (2010). Who Knows Television? Online Access and the Gatekeepers of Knowledge. *Critical Studies in Television*. Autumn 2010, Vol. 5:2, 1-19.

Selkälä, Elina (2016). Yle arkiston aloittamisvuosi? Sähköpostiviesti tekijälle 18.3.2016.

Veyrat-Masson, Isabelle (2012). The Archives of French Television: The Challenge of the 4Cs. *Critical Studies in Television*. Autumn 2012, Vol. 7:2, 36-45.

Tietoa kirjoittajasta

Eeva Savolainen, vastaava luetteloija
Kansallinen audiovisuaalinen instituutti
Radio- ja televisioarkisto
eeva.savolainen@kavi.fi