

*Eeva Savolainen*

# Informaatioalan ammattilaisen roolin muutos metadatan tuotannossa: case Kansallisen audiovisuaalisen instituutin radio- ja tv-tietokanta Ritva

*Eeva Savolainen, Kansallinen audiovisuaalinen instituutti, eeva.savolainen@kavi.fi.*

**M**etadatan tuotantoa sävyttää ristipaine. Yhtäältä metadatan tuotantoa saatetaan pitää toisarvoisena, toisaalta metadatan merkitys kasvaa digitaalisten aineistomäärien lisääntyessä (McCarthy 2014, 1–2, Frederick 2016, 8). Esimerkiksi resurssien löytäminen, sähköisten aineistojen organisointi ja tietokantojen yhteensopivuus ovat riippuvaisia laadukkaasta metadatasta (NISO 2004, 1–2).

Muistiorganisaatiot tekevät yhteistyötä metadatan tuotannossa, ja yleisöäkin pyritään ottamaan prosessiin mukaan. Metadatan tuottamista pidetään kuitenkin edelleen pitkälti informaatioalan ammattilaisten tehtävänä. Mutta tarvitseeko metadatan tuotantoa varjella informaatioalan ulkopuolisilta, toisilla periaatteilla toimivilta tekijöiltä? Millaisia riskejä ”toisten” tuottama metadata tuo tullessaan? Millainen on informaatioalan ammattilaisen rooli, jos metadataa tuottavat myös muut kuin me?

## **Ritvan metadatakokonaisuus**

Radio- ja televisioarkistoissa on jonkin aikaa eletty tilanteessa, jossa metadataa tuottavat myös muut kuin arkistolaiset. Aiemmin kuvailija tuot-

ti metadatan ohjelman lähetyksen jälkeen, nyt erilaiset inhimilliset ja ei-inhimilliset toimijat tuottavat tietoa koko ohjelmanteon ajan (Noordgraaf 2010, 8–9). Yleisradiossa toimittajat ja tuotantokoordinaattorit tuottavat suuren osan ohjelmiin liittyvästä metatiedosta.

Kansallisen audiovisuaalisen instituutin (KAVI) radio- ja tv-arkiston Ritva-tietokanta rakentuu useiden eri toimijoiden tuottamasta metadatasta. Keskeisimmät metadatalähteet ovat suomalaisten mediayhtiöiden ohjelmatiedotusjärjestelmä Venetsia, kanavien ohjelmaoppaat, televisionkatselua ja radionkuuntelua mittaava Finnpanel Oy, suomalaisten radiokanavien Radiot.fi-tietojärjestelmä sekä Yleisradio. Kukaan em. toimijoista Yleisradiota lukuun ottamatta ei tuota metadataa arkistointinäkökulmasta. Venetsian aineisto on suunnattu toimittajille, kanavien ohjelmaoppaat ja Radiot.fi:n tiedot ohjelmien loppukäyttäjille. Finnpanelin toimeksiantajia ovat mediayhtiöt, joita kiinnostavat ohjelmien katsoja- ja kuuntelijaluvut.

## **Riesoja ja hyötyjä**

Monilähteisen metadatakokonaisuuden keskeinen haaste on tietojen tuplaantuminen. Eri läh-

teistä saatuja ohjelmätietoja yhdistellään mahdollisimman paljon automaattisesti, mutta tiettyjen raja-arvojen jälkeen automaattiyhdistely ei toimi, vaan yhdistely on tehtävä käsin. Tietojen tuplaantumista tapahtuu myös ohjelmätietueiden sisällä, mutta niitä ei käsitellä käsin vaan järjestelmä piilottaa ylimääräiset kuvauskentät. Henkilörekisterin tuplatietueita siistitään osittain automaattisesti ja käsin.

Riesana ovat myös metadatan virheet: kahden henkilön tietoja on ahdettu yhteen tietuekorttiin, kirjoitusvirheet rehottavat, kentissä on tietoja jotka eivät sinne kuulu, päällekkäiset kooditiedot aiheuttavat sotkuja. Räikeimmät tiedonhakuja haittaavat virheet pyritään korjaamaan käsityönä.

Automaattisen metadatan kautta ei saada kaikkia tarvittavia tietoja. Tiedot saattavat puuttua kokonaan, jolloin niitä pitää kerätä eri lähteistä. Joitakin tietoja (esim. tuotantomaista) saadaan vain osalta kanavista. Tietueita täydennetään valikoiden, esimerkiksi kotimaisiin sarjoihin lisätään käsin tuotantomaaksi Suomi. Lisäksi tuotetaan metadataa esimerkiksi kotimaisista sarjoista ja kanavista, sillä näitä tietoja ei saada ulkopuolelta.

Tekijöiden ja markkinoijien tuottamat sanalliset ohjelmakuvaukset ovat osuvia, mutta niitä kuorruttaakin usein markkinointijargon, jossa ei ole huomioitu FIAF:n (International Federation of Film Archives) ihannetta tiivistelmätekstien objektiivisuudesta. Arkistolla ei kuitenkaan olisi mahdollisuuksia luoda yksittäisiä kuvauksia jokaisesta ohjelmasta erikseen.

### **Yksintekijästä yhteistekijäksi**

Arkisto toimii Ritva-tietokannassa metadatan tuottajaroolin ohella ulkopuolisen metadatan laadunvalvojana, virheiden korjaajana ja täydentäjänä. Vaikka automaattisessa metadatassa on heikkouksia, on tehokkaampaa korjata valmista metadataa kuin tuottaa kaikki itse. Oma-raisesta metadatan tuottamisesta on siirrytty kohti yhteistekijyyttä, jossa metadataa ei välttämättä aina muokata informaatioalan ihannemuotoon. Metadatan laatuun pyritään vaikuttamaan tekemällä yhteistyötä metadatan tuottajien kanssa. Julia Noordergraafin (2010,

9) toteamus informaatioalan ammattilaisen roolin muutoksesta tuotettavan tiedon portinvartijasta eri toimijoiden tuottaman tiedon toimittajaksi on Ritva-tietokannan kannalta varsin osuva.

Muistiorganisaation tuottama kuvailutieto ei välttämättä edes aina ole käyttäjän kannalta taroituksenmukaisinta (ks. esim. McKee 2011, 166–170). Metadatan tuottamisen yhteydessä puhutaankin yhä useammin joukkoistamisesta (esim. Noordegraaf 2010, 9, Besser 2016, 105). Kuvailun apuna hyödynnettävät käyttäjät voivat olla tavallisten kansalaisten lisäksi esimerkiksi tutkijoita (ks. Lykke & Lund & Skov 2016). Ritvassa joukkoistamisesta on toivottu apua sisällön ja metadatan virheiden kartoittamisessa tarjoamalla jokaisessa tietueessa palautemahdollisuus.

Muistiorganisaatioiden metadataan liittyvä auktoriteetti saattaa vähitellen joutua väistymään laajempaan yhteisymmärrykseen perustuvan metadatakäsityksen tieltä (Noordegraaf 2010, 15). Informaatioalan ammattilaisten ja muiden metadatan tuottajien roolien välinen rajankäynti tulee jatkumaan edelleen. Myös automaattiset sisällönanalyyysivälineet kuten kuvan- ja puheentunnistus tulevat muuttamaan kuvailutyötä. Tärkeä on löytää ja kehittää hyvät käytännöt siihen, miten informaatioalan ammattilaisten ja muiden toimijoiden osaaminen voidaan yhdistää mahdollisimman tehokkaasti.

### **Lähteet**

- Besser, Howard (2016). 25 years of digital archives of visual materials: what we've done, what we've learned, and what challenges remain. *Visual Studies*, 31:2, 95–108.
- Frederick, Donna E. (2016). *Managing ebook metadata in academic libraries: taming the tiger*. Waltham, MA: Chandos Publishing, 2016.
- Lykke, Marianne & Lund, Haakon & Skov, Mette (2016). User-Driven CHAOS: Tags and Annotations in Radio Broadcast Research. *Knowledge Organization*. 43:2, 73–85.

McCarthy, Kate (2014). *Metadata Quality Control*. Dublin: Royal Irish Academy. DOI: 10.3318/DRI.2015.1 (13.7.2016)

McKee, Alan (2011). YouTube versus the National Film and Sound Archive: Which is the More Useful Resource for Historians of Australian Television? *Television & New Media*. March 2011, 12:2, 154–173.

NISO, National Information Standards Organization (2004). *Understanding Metadata*. National Information Standards Organization. (13.7.2016)

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