



# RATKAISUJA DIGITAALISTEN AINEISTOJEN KÄYTETTÄVYYDEN PARANTAMISEEN

Miia Kosonen (toim.)



Kaakkois-Suomen  
ammattikorkeakoulu

Miia Kosonen (toim.)

# RATKAISUJA DIGITAALISTEN AINEISTOJEN KÄYTETTÄVYYDEN PARANTAMISEEN



Euroopan unioni  
Euroopan aluekehitysrahasto

Vipuvoimaa  
EU:lta  
2014–2020

Digitalia  
Digitaalisen tiedonhallinnan  
tutkimus- ja kehittämiskeskus



XAMK KEHITTÄÄ 80

KAAKKOIS-SUOMEN AMMATTIKORKEAKOULU  
MIKKELI 2019

© Tekijät ja Kaakkois-Suomen ammattikorkeakoulu

Kannen kuva: Mainostoimisto Ilme Oy

Taitto ja paino: Grano Oy

ISBN: 978-952-344-181-1 (nid.)

ISBN: 978-952-344-182-8 (PDF)

ISSN: 2489-2467 (nid.)

ISSN 2489-3102 (verkkajulkaisu)

[julkaisut@xamk.fi](mailto:julkaisut@xamk.fi)

# DIGITAALISET AINEISTOT KÄYTTÖÖN!

Juha Sipilän johtaman hallituksen kantava teema oli digitalisaatio. Hallituskauden tavoitteena oli kehittää käyttäjälähtöiset, tuottavuutta ja tuloksellisuutta nostavat yhden luukun digitaaliset julkiset palvelut. Yksi kärkihankkeista oli *Digitalisoidaan julkiset palvelut*, jonka toteuttamisesta vastasi valtiovarainministeriö. Kärkihankkeen avulla toteutettiin useita erillisiä hankkeita. Käyttöön otettiin muun muassa suomi.fi –palvelut.

Valtioneuvosto teki 21.6.2017 periaatepäätöksen asiakirjallisen aineiston digitoinnista ja arkistoinnista vain sähköisenä. Tämän päätöksen jälkeen Kansallisarkistossa luotiin etene-  
missuunnitelma. Se tähtää siihen, että edelleen viranomaisissa oleva arkistoitava analoginen aineisto muunnetaan digitaaliseen muotoon. Aineistoa on paljon – 160 kilometriä! Aineistoa on pääsääntöisesti tuotettu 1970-luvulla tai myöhemmin, joten suurimmalle osalle aineistoa voidaan suorittaa automaattinen sisällöntunnistus.

Mahdollisuuksia on paljon, kun tieto ei ole enää pelkkä bittikarttakuva. Aineistoista voidaan louhia tietoja, tuottaa parempia hakuja, helpottaa löydettävyyttä ja tarkentaa kuvailua. Puhetta tekoälystä, algoritmeista ja koneoppimisesta on paljon, mutta osajia ja tuloksia vähemmän. Erilaisia projekteja ja ryhmiä onneksi on alkanut syntyä. Vuonna 2015 perustettu Digitalia lienee yksi vanhimmista, mikä kertoo kehityksen vauhdista.

Tämä artikkelikokoelma esittelee Digitaaliset aineistot käyttöön -hankkeen tuloksia. Kansallisarkisto ja Digitalia ovat tehneet hankkeessa yhteistyötä, jonka tuloksena Kansallisarkisto on organisaationa viisaampi liittyen sisällönanalysointiin. Tekemämme seulontapäätökset ovat helpommin saatavilla asiakkaille. Yhteistyö luo pohjaa sille työlle, jota on ryhdytty Kansallisarkistossakin viemään eteenpäin – ja joka osaltaan toteuttaa jo edesmenneen hallituksen ohjelmaa.

Helsingissä 6.5.2019

*Reko Etelävuori*

ylitarkastaja, sisällönanalysointiprojektin projektipäällikkö, Kansallisarkisto

## KIRJOITTAJAT

**Torsti Hyyryläinen**, HT, tutkimusjohtaja  
Helsingin yliopisto, Ruralia-instituutti

**Teemu Hänninen**, it-tradenomi, it-asiantuntija  
Kaakkois-Suomen ammattikorkeakoulu

**Anssi Jääskeläinen**, TkT, tki-asiantuntija  
Kaakkois-Suomen ammattikorkeakoulu

**Kimmo Kettunen**, FT, tutkimuskoordinaattori  
Kansalliskirjasto

**Mika Koistinen**, DI, tietojärjestelmäasiantuntija  
Kansalliskirjasto

**Miia Kosonen**, KTT, tki-asiantuntija  
Kaakkois-Suomen ammattikorkeakoulu

**Johanna Lilja**, FT, palvelujohtaja  
Kansalliskirjasto

**Erno Liukkonen**, FM, tietojärjestelmäasiantuntija  
Kansalliskirjasto

**Juuso Muttilainen**, LuK, projektityöntekijä  
Helsingin yliopisto, Ruralia-instituutti

**Tuula Pääkkönen**, FM, tietojärjestelmäasiantuntija  
Kansalliskirjasto

**Manu Rantanen**, FM, projektipäällikkö  
Helsingin yliopisto, Ruralia-instituutti

**Teemu Ruokolainen**, TkT, tutkijatohtori  
Kansalliskirjasto

**Tuomo Räisänen**, FT, it-asiantuntija  
Kaakkois-Suomen ammattikorkeakoulu

**Noora Talsi**, YTT, tutkimusjohtaja  
Kaakkois-Suomen ammattikorkeakoulu

## SISÄLTÖ

JOHDANTO .....	6
Noora Talsi	
KANSALLISKIRJASTON HISTORIALLISEN LEHTIKOKOELMAN KEHITTÄMISTYÖ .....	11
Kimmo Kettunen, Mika Koistinen, Teemu Ruokolainen, Erno Liukkonen, Tuula Pääkkönen	
AUTOMATISOINNILLA ETUJA ASIAKIRJAHALLINTAAN .....	21
Anssi Jääskeläinen, Tuomo Räisänen	
3D-PELIMOOTTORIN HYÖDYNTÄMINEN HISTORIALLISTEN TAPAHTUMIEN VISUALISOINNISSA .....	29
Anssi Jääskeläinen	
AVOIMILLA OHJEISTUKSILLA JA OPPAILLA TUETAAN DIGITAALISEN TIEDON PITKÄAIKAISSÄILYTTÄMISTÄ .....	34
Teemu Hänninen	
DIGITAALISTEN AINEISTOJEN SISÄLTÖKÄYTTÖTOIVEITA TUTKIJOILTA .....	40
Tuula Pääkkönen, Johanna Lilja	
HISTORIALLISET LEHDET YRITYKSIÄ INSPIROIVANA TARINA-AINEISTONA .....	48
Torsti Hyyryläinen, Manu Rantanen, Juuso Muttilainen	
TIETO HYÖTYKÄYTTÖÖN TIEDOLLA JOHTAMISTA KEHITTÄMÄLLÄ .....	58
Miia Kosonen	
DIGITALIA LUO MAHDOLLISUUKSIA .....	64
Noora Talsi, Miia Kosonen	

# JOHDANTO

Noora Talsi, YTT, tutkimusjohtaja, Kaakkois-Suomen ammattikorkeakoulu

Digitaalisen tiedon määrä lisääntyy maailmassa valtavaa vauhtia ja sen merkitys päätöksenteon perustana kasvaa. Digitaalista tietoa syntyy lähes kaikessa toiminnassa, puhuttiinpa sitten yksilöistä, yhteisöistä tai yhteiskunnasta. Digitalian tavoitteena on näiden tietojen parempi käytettävyys ja saatavuus.

Digitaalisen tiedonhallinnan tutkimus- ja kehittämiskeskus Digitalia perustettiin vuonna 2015 Kaakkois-Suomen ammattikorkeakoulun, Kansalliskirjaston ja Helsingin yliopiston yhteiseksi verkostomaiseksi osaamiskeskukseksi ja yhteistyömalliksi. Digitalian toiminnan ydintä on digitaalisen tiedon koko elinkaaren hallinnan prosessi: tiedon keruu, tallentaminen, analysointi, käyttöön saattaminen, arkistointi ja pitkäaikaissäilytys.

## UUTTA OSAAMISTA AINEISTOJEN KÄSITTELYYN

Digitaaliset aineistot käyttöön –hankkeen tavoitteena on ollut aineistojen käytön ja käytettävyyden edistäminen muun muassa aineistoja visualisoimalla, anonymisoimalla ja haettavuutta parantamalla. Hanke on siten luonut uutta digitaalisten tietomassojen osaamista.

Tässä artikkelikokoelmassa esitellään Digitaaliset aineistot käyttöön –hankkeessa tehtyjä tuloksia. Samalla Digitalian taustaorganisaatioissa Kaakkois-Suomen ammattikorkeakoulussa, Helsingin yliopiston Ruralia-instituutissa ja Kansalliskirjastossa tehty kehittämistyö ja aineistot tulevat lukijalle tutummiksi. Tuloksia voivat hyödyntää digitaalisten aineistojen säilyttäjät, tutkijat ja käyttäjät.

Ensimmäiset kolme artikkelia käsittelevät digitaalisessa muodossa olevan tiedon käytettävyyden parantamista. Artikkelikokoelman ensimmäinen artikkeli käsittelee Kansalliskirjaston historiallisten lehtien, kirjojen ja pienpainatteiden kokoelman kehittämistä ja käytettävyyden parantamista. Automatisoinnilla etuja asiakirjahallintaan –artikkelissa pureudutaan PDF-muotoisten aineistojen käytön parantamiseen ja niihin liittyviin haasteisiin. Kolmannessa artikkelissa on testattu arkistoaineiston käyttäjäkokemuksen kehittämistä pelimootorin avulla.

Digitaaliset aineistot käyttöön –hankkeessa selvitettiin aineistojen käyttömahdollisuuksien laajentamista niin yksittäisten kansalaisten kuin erityyppisten organisaatioiden ja yritysten tarpeisiin. Neljäs artikkeli esittelee erityisesti pienten yhteisöjen ja kansalaisten käyttöön tarkoitettujen kansantajuisen aineistojen digitointioppaan. Viidennessä artikkelissa katse käänn-

tään käyttäjiin itseensä ja käydään läpi digitaalisten lehtiaineistojen käyttöoikeuksiin liittyviä palautteita ja toiveita eri käyttäjäryhmiltä. Kuudennessa artikkelissa kuvataan digitaalisten lehtiaineistojen sisällön louhintaa yritysten käyttöön ja hyödynnettäväksi esimerkiksi yrityksen tarinan rakentamisessa. Kokoelman päättävässä seitsemännessä artikkelissa tarkastelun kohteena ovat yritysten edellytykset tiedon hyödyntämiseen eli niiden valmiudet johtaa tiedolla.

## LIIKETOIMINTASELVITYKSEN TULOKSIA

Xamkilla on aiemmin kehitetty Kansalaisarkistoa, jonka tavoitteena on ollut tarjota helppokäyttöinen arkistoratkaisu kansalaisten käyttöön. Samalla se on tarjonnut Digitalialle alustan, jolla sähköistä arkistointia voidaan pilotoida. Osana Kansalaisarkiston kehittämissuunnitelmaa Digitaliassa teetettiin hankkeen aikana liiketoimintaselvitys. Selvityksen tärkein menetelmällinen viitekehys rakentui palvelumuotoiluprosessin ympärille. Toisena viitekehysnä oli Lean Startup ja liiketoimintasuunnitelman laatiminen nopeilla kokeiluilla ja mittaamisella.

Liiketoimintaselvitys kilpailutettiin ja sen tekijäksi valittiin Osaango Oy. He käyttivät liiketoimintaselvityksen pohjana asiakaslähtöisyyden mallia (Korkman & Arantola 2009, 69). Se soveltui projektiin hyvin, koska Kansalaisarkiston osalta tuotekehitystä oli asiakkaan sijaan tehty tutkimus edellä. Tavoitteena oli tavallisten käytäntöjen ja sosiaalisen kontekstin löytäminen arkistointiin osana ihmisten ja yhteisöjen elinkaarta (emt. 8). Erityisen hyvin malli soveltui Osaango Oy:n mukaan kohdealueeseen eli arkistointiin. Kaikki yritykset ja yhteisöt joutuvat lakisääteisesti huolehtimaan arkistoinnista, mutta usein se on niille hankala käytäntö. Sitä ei ole kiinnostusta opetella, sitä koetaan tarvittavan harvoin tai sen tarvetta ei edes tiedosteta kunnolla. Kaupallistamisselvityksessä selvisi myös, että suurin haaste yhdistyksillä ja seuroilla on tiedon ja ajan puute. (Niinioja 2018)

Liiketoimintasuunnitelmassa haluttiin selvittää, löytyykö jo kehitetyn kaltaisille tuotteille ylipäättään markkinaa ja mitä ominaisuuksia tuotteissa pitäisi olla. Optimaalisen tuotteen löytäminen saattaa olla vaikeaa, ja tarpeet elävät jatkuvasti uusien markkinoille tulevien innovaatioiden ja asiakkaiden kulutustottumusten muuttumisen vuoksi. Innovaatioiden kehittämisessä tuleekin varautua siihen, että tarve – tässä tapauksessa tarve arkistoida ja tapa toteuttaa arkistointi – vaihtelee ajan saatossa ja ympäristön mukaan. Tietojen säilytystä ja dokumentteja tulisi tutkia niiden historian kautta, jotta ymmärretään aineistojen taustalla oleva tarve ja tuleva kehityskulku. (Niinioja 2018)

Liiketoimintamallista muodostui erilainen kuin alussa oletettiin. Selvityksen voidaan siis hyvällä syyllä sanoa tuottaneen arvokasta tietoa. Digitaaliset aineistot käyttöön -hankkeen alussa Kansalaisarkisto nähtiin tärkeimpänä kaupallistettavana tuotteena ja sidosryhmien sanoin ”Digitalian lippulaivana” Kansalliskirjaston digitoitujen lehtien kokoelman ohella. Hankkeen aikana päädyttiin malliin, jossa Kansalaisarkisto on taustakonsepti ja arkkitehtuurisuositus, johon liittyvää osaamista jaetaan oppaiden, konsultoinnin ja räätälöityjen ratkaisujen kautta.



Liiketoimintaselvityksessä huomio kiinnittyi Kansalaisarkiston raskaaseen ja kalliiseen infrastruktuuriin, todellisen asiakassegmentin kapeuteen sekä infraan nähden riittämättömään tulovirtaan asiakkailta. MyData -periaatteiden mukaisesti Digitalia on pyrkinyt edistämään kansalaisten oikeuksia hallita ja jakaa omia tietojaan, mutta liiketoimintaselvityksen osana tehdyissä taustahaastatteluissa tiivistyi osuvasti tähän liittyvä ristiriita: *”Henkilökohtainen arkisto on taideteos, ei aktiivikäytössä oleva tietopankki.”* Jatkokehityksen kannalta Kansalaisarkiston haasteeksi muodostui myös raskas rakenne, joka edellyttää käyttäjältään vähintään perustason arkistointiosaamista. Jo tuotetun ratkaisun päälle osoittautui vaikeaksi kehittää kansalaiskäyttöön soveltuvaa arkistoa, jota kuka tahansa osaisi intuitiivisesti käyttää, ja tietojen syöttäminen ja hallinta olisi mahdollisimman helppoa. Monille myös riittävät pilvipalvelut tai ulkoiset kovalevyt tietojen säilytykseen, vaikka ne eivät varsinaista arkistointia tai pitkäaikaissäilytystä edustakaan.

## YHDISTYKSET TARVITSEVAT OPASTUSTA ARKISTOINTIIN

Tehdyn selvityksen mukaan pienten arkistotoimijoiden kannalta tarpeellisinta on kouluttaminen ja tiedon levittäminen. Niinpä suosituksiksi muodostui toteuttaa tekniset, juridiset ja toiminnalliset lähtökohdat huomioiva opas suunnattuna erityisesti yhdistyksille tai muille sellaisille toimijoille, joilla ei ole käytettävissä arkistohallinnan ammattilaista, mutta joilla on lakisääteinen tarve säilyttää tietoa. Nykyiset ilmaiset ja maksulliset oppaat ovat vanhentuneita ja epäselviä, tieteelliset artikkelit puolestaan liian vaikeita ymmärtää. Tuotettavan tiedon tulisi olla selkokielistä ja nopeasti omaksuttavaa, samaan aikaan kun sen pohjautuu tutkimukseen. (Niinioja 2018)

Opastamiseen, uuden tiedon jakamiseen ja ajattelun herättämiseen soveltuvat kuviin, info-graafeihin, videoihin ja blogeihin perustuvat nykyaikaiset materiaalit, jotka mahdollistavat mikro-oppimisen. Tämä istuu hyvin tutkimusorganisaation imagoon tiedon jalostajana ja jakajana. Mikro-oppimisen konkreettisina välineinä toimivat tietosisällöt esimerkiksi sosiaalisen median kampanjoissa, yhteistyökumppaneiden sivustoilla ja sähköpostiviestinnässä. (Niinioja 2018)

Liiketoimintaselvityksen suositusten pohjalta Digitaliassa onkin koottu erityisesti kansalaisille ja pienyhteisöille sopiva Digitointiopas (Hänninen 2019, tässä kokoelmassa). Myös jatkossa Digitaliassa tullaan tuottamaan avoimesti saatavilla olevia digitaalisen tiedonhallinnan ja arkistoinnin opas- ja ohjemateriaaleja. Ne lisäävät sidosryhmien ymmärrystä tarjolla olevista ratkaisuista, tukevat niiden käyttöönottoa ja pitkällä aikajänteellä edistävät digitaalisen tiedon säilymistä ja hyödyntämistä pitkälle tulevaisuuteen.

## **MYÖS SÄHKÖPOSTIEN ARKISTOINTIA ON KEHITETTY ETEENPÄIN**

Digitaalisessa muodossa olevan aineiston arkistointi ei ole ongelmaton. Digitalia on tehnyt yhteistyötä Päivälehdien arkiston kanssa sähköpostiaineiston arkistoinnissa. Haasteena olivat Helsingin Sanomien entisten päätoimittajien sähköpostit, jotka oli luovutettu arkistolle. Jos on yrittänyt avata vanhoja sähköposteja nykyaikaisilla ohjelmistoilla – tai toisinpäin – tietää, kuinka hankalaa viestejä on saada auki, puhumattakaan päätoimittajilta jääneistä kymmenien tuhansien sähköpostien aineistoista. Jääskeläinen (2017) kehitti menetelmän, jolla koko sähköpostilaatikko voidaan arkistoida niin, että alkuperäinen viesti- ja kansiorakenne säilyy ja kaikki liitteet on arkistoitu yhdessä kyseisen sähköpostin kanssa. Näin päätoimittajien kirjeenvaihto säilyy tuleville sukupolville tutkijoiden, toimittajien ja historiankirjoittajien käyttöä varten.

Sähköpostien arkistoinnin kehitystä on jatkettu eteenpäin muidenkin toimijoiden tarpeisiin. Sähköpostien automaattinen arkistointi tukee käytännössä kaikkia käytössä olevia sähköpostiformaatteja esimerkiksi Outlook, Gmail ja Hotmail. Työkalua on testattu yhdessä Disec Oy:n kanssa. Digitalia on kehittänyt rajapinnan, jonka kautta sähköpostityökalu on käytettävissä. API tullaan julkaisemaan Digitalian verkkosivuilla.

Etelä-Savon Maakuntaliiton EAKR-rahoituksella toimineen hankkeemme tuloksena useilla eri kohderyhmillä – arkistot, kirjastot, tutkimusorganisaatiot, yritykset ja kansalaiset – on paremmat mahdollisuudet käyttää tietoa ja luoda siitä uutta arvoa. Digitaaliset aineistot on saatettu laajempaan käyttöön ja käyttäjillä on paremmat mahdollisuudet selata, hakea ja jakaa aineistoja. Toivotamme lukijoille innostavia hetkiä artikkelikokoelman ja digitaalisten aineistojen aarreaitan parissa!

## LÄHTEET

Jääskeläinen, A. & Kokkonen, T. 2017. Sähköpostien ihanuus ja kurjuus. Teoksessa Kosonen, M. (toim.) Digitaalinen tieto haltuun. Mikkeli: Xamk kehittää 11, Kaakkois-Suomen ammattikorkeakoulu, 48-57.

Korkman, O. & Arantola, H. 2009. Arki, Eväitä uuteen asiakaslähtöisyyteen. Helsinki: WSOYpro.

Niinioja, M. 2018. Liiketoimintasuunnitelma Digitalialle. Projektin loppuraportti.

# KANSALLISKIRJASTON HISTORIALLISEN LEHTI- KOKOELMAN KEHITTÄMISTYÖ

Kimmo Kettunen, FT, tutkimuskoordinaattori & Mika Koistinen, DI,  
tietojärjestelmäasiantuntija & Teemu Ruokolainen, TKT, tutkijatohtori &  
Erno Liukkonen, FM, tietojärjestelmäasiantuntija & Tuula Pääkkönen, FM,  
tietojärjestelmäasiantuntija, Kansalliskirjasto

Kansalliskirjaston Mikkelin digitointi- ja konservointikeskus on digitoinut Suomessa julkaistuja sanoma- ja aikakauslehtiä sekä pienpainatteita kattavasti vuodesta 1998 alkaen. Koko aineisto on käytettävissä Kansalliskirjaston verkkopalvelussa [digi.kansalliskirjasto.fi](http://digi.kansalliskirjasto.fi). Aineiston määrä on kasvanut vuosien aikana tasaisesti, samoin aineiston käyttäjien määrä. Lehtiaineistoa on verkkopalvelussa vapaasti saatavilla vuosilta 1771–1929. Koko aineisto vuosilta 1771–1910 on saatavilla myös avoimen datan jakelupakettina verkkopalvelun sivuilta (Pääkkönen ym. 2016). Vuonna 2018 aineistoon lisättiin 2085 kirjan klassikkokirjasto kotimaista kirjallisuutta.

Kansalliskirjasto ilmoittaa verkkosivuillaan digitoitujen sivujen kokonaismääräksi vuoden 2019 huhtikuun lopulla 16 179 109. Luku sisältää sanoma- ja aikakauslehdet, kirjat sekä pienpainatteet. Vapaasti käytettäviä lehtiä on noin 7,59 miljoonaa sivua: niistä noin 4,08 miljoonaa sivua on sanomalehtiä, 3,38 miljoonaa sivua aikakauslehtiä ja 129 343 sivua erilaisia pienpainatteita. Kuva 1 esittää kokoelman käyttöliittymän aloitussivua.

The screenshot shows the homepage of the digital archive. At the top, there is a dark blue header with the logo of the National Library of Finland (Kansalliskirjasto) and the text "DIGITAALISET AINEISTOT BETA". There are buttons for "Palautte" (Feedback) and "Kirjaudu" (Log in). Below the header is a navigation menu with links for "HAKU" (Search), "KOKOELMAT" (Collections), "LEHDIT" (Newspapers), "PÄIVÄN LEHDET" (Today's Newspapers), and "LEIKKEET" (Clippings). A search bar is present with the placeholder text "Hae digitoituista aineistoista". Below the search bar are several category thumbnails: "SANOMALEHDET" (Newspapers), "AIKAKAUSLEHDET" (Magazines), "PIENPAINATTEET" (Microforms), "KIRJAT" (Books), "KÄSIKIRJOITUKSET" (Manuscripts), "KARTAT" (Maps), and "NUOTIT" (Scores). Each thumbnail has a representative image and a title. At the bottom, there is a footer with the text "Tutustu historiaan helposti laajojen digitaalisten aineistojen avulla!" and the date "03.04.2019". There is also a count of "16 179 109 sivua aineistoja" and logos for the European Union and "Vipuvoimaa EU:lta".

Kuva 1. *Digi.kansalliskirjasto.fi:n* etusivu

Kansalliskirjaston digitoitu lehtikokoelma on osa maailmanlaajuisista laajenevaa historiallisten lehtien digitoitua tarjontaa. Europeana-projekti arvioi vuonna 2012 Euroopassa olevan digitoituja sanomalehtiä noin 129 miljoonaa sivua ja noin 24 000 nimikettä (Dunning 2012). Varovainen arvio digitoitujen lehtinimikkeiden määrästä muutaman vuoden takaa maailmanlaajuisesti on yli 45 000. Suurin osa aineistosta on digitoitu Euroopassa ja Yhdysvalloissa. (The ”State of the Art” 2015). Pohjoismaissa oli digitoitu arviolta 84 miljoonaa sivua vuoden 2018 alkuun mennessä (Pääkkönen ym. 2018).

Digitaaliset historialliset lehtiaineistot ovat tärkeä osa kansakunnan muistia. Aineistot muodostavat digitaalisen lehtikirjaston ytimen, mutta ilman jatkuvaa kehitystyötä palvelu ei vastaa riittävän hyvin käyttäjien vaihteleviin käyttötarpeisiin, olipa käyttö sitten kansalaisten vapaa-ajalla tapahtuvaa tai tutkimusta. Digitaaliset aineistot käyttöön –hankkeessa Kansalliskirjasto kehitti ja tutki digi.kansalliskirjasto.fi –palvelua ja sen aineistoja eri tavoin. Sanomalehtiaineiston suomenkielisen osan optisen luvun tasoa pyrittiin kehittämään uudella optisen luvun työnkululla. Sanomalehtiaineistosta tehtiin uusi nimien tunnistamisen koe- ja opetusaineisto, jolla tutkittiin erityisesti henkilön- ja paikannimien tunnistamisen mahdollisuutta kokoelmassa. Uuden Suomettaren aineistolla valmistui artikkelien eristämisen opetus- ja testiaineisto, jolla arvioitiin artikkelien eristämisen mahdollisuuksia ja laatua. Lehtiaineiston sisältämille kuville tehtiin luokituskokeita niiden kuvatyypeistä. Palvelun käyttöliittymään tehtiin erilaisia käyttäjiä palvelevia muutoksia. Näistä kaikista annetaan lyhyt katsaus tässä artikkelissa. Lopuksi kerrotaan lyhyesti hankkeen aikana tehdystä tutkimusyhteistyöstä.

## UUSI OPTISEN LUVUN TYÖNKULKU

Digitoidut lehtiaineistot tuotetaan kuvaamalla lehtien sivut ja tunnistamalla sivujen sisältö optisella merkkien tunnistuksella (OCR, Optical Character Recognition). Optinen merkin-tunnistusohjelma vastaa viime kädessä siitä, miten hyvin kuvatus sivun teksti tunnistuu. Koska kirjaston käytössä ollut optisen luvun ohjelmisto alkaa olla vanha eikä sitä ole mahdollista päivittää lisenssien kalleuden vuoksi, olemme tehneet hankkeen aikana työtä uuden avoimeen lähdekoodiin perustuvan Tesseract-ohjelmiston<sup>1</sup> käyttöön saattamisessa.

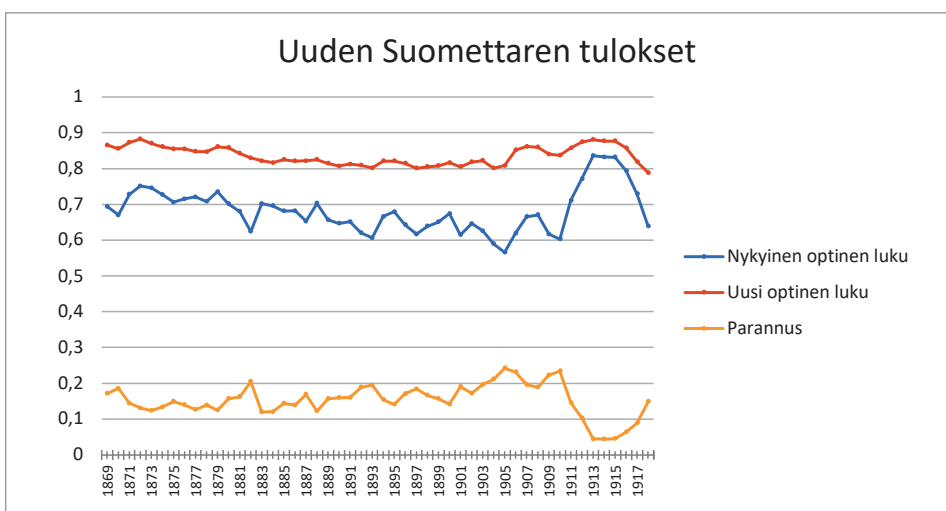
Aineiston uutta optista lukemista varten luotiin ensin noin 500 000 sanan otos, josta tehtiin ihmistyönä tarkistettu mahdollisimman virheetön versio. Tätä aineistoa voidaan käyttää tekstien laadun parantumisen arvioimisen testiaineistona.

Eriyisesti sanomalehdet on painettu Suomessa vielä 1900-luvun alkuvuosikymmeniin saakka pääosin fraktuura-kirjasimella. Kirjasin on tunnetusti hyvin hankala tulkittava optisille lukuohjelmille (Holley 2009; Piotrowski 2012). Tesseractia varten luotiin hankkeessa malli suomen kielen fraktuura-kirjasimelle ja samalla parannettiin ohjelman kykyä käsitellä sanomalehtien digitoituja sivukuvatiedostoja.

---

<sup>1</sup> <https://github.com/tesseract-ocr>

Pitkän kehitystyön jälkeen Tesseractilla tehdyssä koeaineiston merkintunnistuksessa sanojen virheprosentti saatiin laskemaan 6.56:een alkuperäisestä 25.30 %:sta, mikä on 18.74 %:n parannus alkuperäiseen (Kettunen & Koistinen 2019). Tähän mennessä koko lehtiaineistosta on tuotettu uusi optinen luku Uuden Suomettaren 86 068 sivulle, joilla on noin 307 miljoonaa sanaa. Tässä aineistossa saavutettiin keskimäärin 15.3 %:n parannus sanojen tunnistettavuuteen automaattisella morfologisella analyysiohjelmalla. Parannettua optista lukua jatketaan lehdille niiden käytön suosittuusjärjestyksessä, ja parempilaatuista lehtiaineistoa sisällytetään Digin käyttöliittymään. Optisen luvun kehittämistä varten luotu testiaineisto on saatavilla [digi.kansalliskirjasto.fi](https://digi.kansalliskirjasto.fi) –sivuilla avoimena aineistona<sup>2</sup>. Kuvassa 2 näytetään sanojen tunnistettavuuden parantuminen prosentteina uudella optisella luvulla nykyiseen verrattuna. Luvut ovat automaattisen morfologisen tunnistusohjelman antamia arvioita sanojen tunnistettavuudessa aineistossa.



Kuva 2. Uuden optisen luvun tulokset Uuden Suomettaren aineistolla 1869-1918

## NIMIEN ERISTÄMINEN LEHTIAINEISTOSTA

Ihmisten, paikkojen ja erilaisten organisaatioiden nimiä pidetään yleisesti tärkeinä erilaisissa historiallisissa aineistoissa. Aineistojen käyttäjät sijoittavat monesti myös itsensä aineiston maailmankartalle sosiaalisesti ja maantieteellisesti nimien kautta. Tiedonhaun tutkimuksesta tiedetään, että historialliseen tekstiaineistoon tehdään paljon hakuja erilaisia nimiä käyttäen. Niinpä nimiä voidaan käyttää myös auttamaan aineiston selailukäyttöä eristämällä tekstiaineistosta nimiä ja tekemällä niiden avulla hakemistoja aineistoon (Neudecker ym. 2014).

Kansalliskirjastossa on luotu hankkeen aikana laaja nimien eristämisen testi- ja opetusaineisto. Sen avulla opetettiin nimien tunnistamisessa yleisesti käytetty tilastolliseen koneop-

<sup>2</sup> <https://digi.kansalliskirjasto.fi/opendata/submit>

pimiseen perustuva Stanford NER –ohjelmisto<sup>3</sup> tunnistamaan henkilöiden ja paikkojen nimiä 1800-luvun ja 1900-luvun alun digitoidussa sanomalehtiaineistossa (Ruokolainen & Kettunen 2018). Saavutettuja tuloksia voi pitää melko hyvinä: ihmisten nimistä tunnistettiin koeaineistoissa 71 %, paikkojen nimistä 79 %, kun käytettiin laadultaan parannettua optisen luvun tekstiaineistoa.

Nimien käyttömahdollisuuksia lehtikokoelmassa on kuvattu tarkemmin lähteissä Kettunen ym. (2016) sekä Kettunen & Ruokolainen (2017). Hyvän käsityksen nimien käytöstä digitoitun sanomalehtiarkiston käyttöliittymässä antaa esimerkiksi italialaisen La Stampan historiallinen kokoelma. Siinä nimiä käytetään artikkelien selauksen tarkennuksessa sen jälkeen, kun kokoelmaan on tehty ensin haku jollain hakusanalla. Olemme toteuttaneet Digin käyttöliittymään Uuden Suomettaren aineistolle ihmisten ja paikkojen tunnistuksen ja merkinnän.

Teimme myös yhteistyötä Aalto-yliopiston tutkijoiden kanssa nimien linkityksessä tietokantoihin pienen koeaineiston avulla. Tulosten mukaan linkitys onnistuu melko hyvin, mutta virheellisesti kirjoitettujen nimien löytäminen tietokannoista ei onnistu, vaikka nimien tunnistamisen ohjelma olisikin ne löytänyt ja merkinnyt tekstiin (La Mela ym. 2019).

## ARTIKKELIEN ERISTÄMINEN AINEISTOSTA

Kansalliskirjaston lehtiaineiston nykyinen esitysmuoto perustuu kokonaisuun digitoituihin painettuihin sivuihin. Aineistoon tehtävän tiedonhaun ja muun käytön kannalta tämä ei ole paras mahdollinen ratkaisu, ja siksi Digitaaliset aineistot käyttöön –hankkeessa yhtenä työnä on ollut artikkelien eristäminen digitoiduilta sivuilta. Artikkelit ovat käyttäjän kannalta lehtien perustava informatiivinen sisältö, ja mahdollisuus kohdistaa tekstihaut suoraan artikkeleihin parantaisi digitoitujen lehtien käytettävyyttä. Myös kaikenlainen muu aineiston poiminta ja käsittely hyötyisi artikkelien eristämisestä.

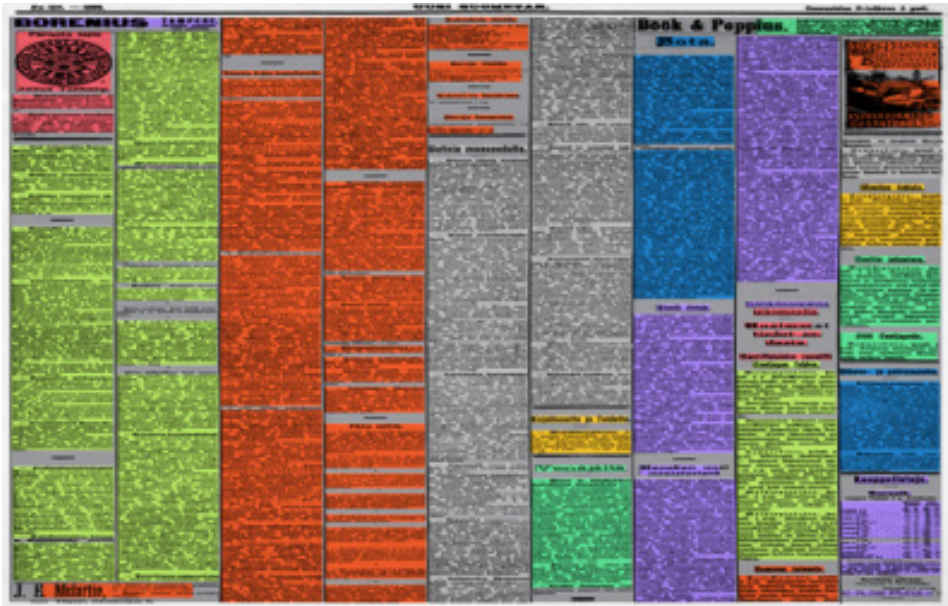
Artikkelien eristämiseen soveltuvien ohjelmistojen ja ratkaisujen etsinnässä teimme työtä jo aiemmin. Saatavilla olevien ohjelmistojen selvitystyön jälkeen päädyimme käyttämään Rouenin yliopiston LITIS-laboratoriossa tehtyä PIVAJ-ohjelmistoa artikkelien eristämässä (Hebert ym. 2014). Hankimme ohjelmistosta lisenssin, ja saimme Rouenin yliopiston tutkijoilta apua ohjelmiston käytössä aineistojemme kanssa.

Päädyimme tekemään työtä artikkelien eristämässä Uuden Suomettaren aineistolla. Lehteä ilmestyi tällä nimellä vuosina 1869–1918 yhteensä 86 068 sivua. Lehdessä oli aluksi kolme palstaa, mutta palstojen määrä lisääntyi enimmillään yhdeksään, josta se palasi takaisin kuu-

---

<sup>3</sup> <https://nlp.stanford.edu/software/CRF-NER.html>

teen palstaan. Palstojen määrän vaihtelun lisäksi artikkelien eristämiseksi tuottavat ongelmia muun muassa mainosten lisääntyminen sekä erilaiset muut sivujen ulkoasun muutokset. Kuvassa 3 on PIVAJ'n tuottama visuaalinen mallisivu Uuden Suomettaren sivun sisällön jakautumisesta eri artikkeleihin. Kutakin ohjelman tunnistamaa artikkelia vastaa oma väri.

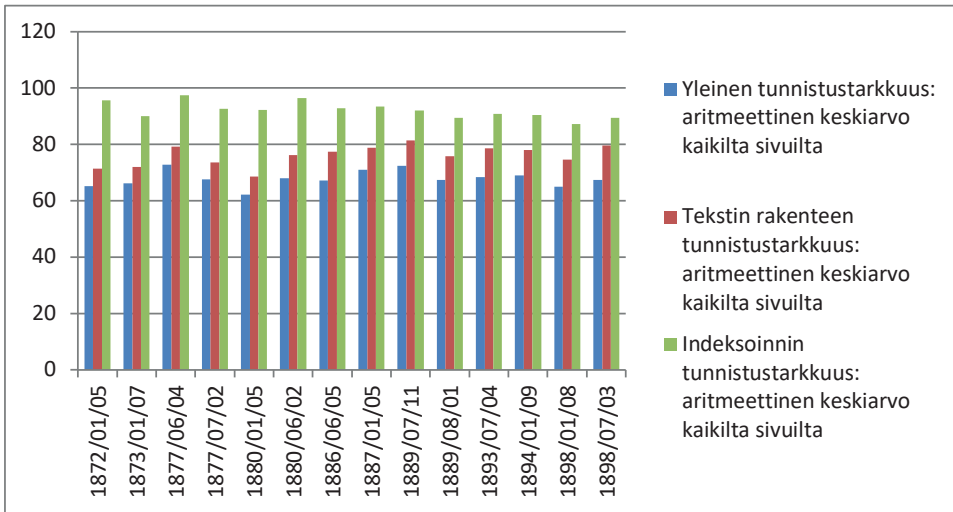


*Kuva 3. PIVAJ'n erottelemat artikkelit Uuden Suomettaren sivulla*

Teimme projektissa PIVAJ'n opettamista ja artikkelien erottamisen laadun arviointia varten opetus- ja koearineiston Uuden Suomettaren sivuista vuosilta 1869–1898. Aineistoon luotiin sivujen rakenne merkitsemällä käsin artikkelien rajat, otsikot jne. Aineistoon tuli yhteensä 224 sivua, joista 168 sivua käytettiin PIVAJ'n opettamiseen ja 56 sivua tulosten arviointiin (Kettunen ym. 2019).

Saavutettuja tuloksia voi pitää kohtuullisina. Arvioimme tuloksia kolmella eri kriteerillä käyttäen ICDAR-konferenssissa käytettyä Salfordin yliopiston PRImA-tutkimuslaboratorion ohjelmistoa (Clausner ym. 2011). Kuvassa 4 on esitetty saavutetut tulokset koearineiston 14 numeron 56 sivulla (Kettunen ym. 2019).





*Kuva 4. PIVAJ'n artikkelien erotuksen tarkkuus koeaineistossa*

Keskimäärin eri arviointitapojen saavuttamat tunnistustarkkuudet koko koeaineistossa olivat 67,9, 76,1 ja 92,2.

Aiomme jatkaa artikkelien eristämistä Uuden Suomettaren aineistolla. Saatuja artikkeleita voidaan liittää osaksi Digin käyttöliittymää käyttämällä esitysjärjestelmässä olevaa leiketointia. PIVAJ automatisoi leikkeiden tekemisen, ja samalla saamme leikkeistä myös niiden tekstisisällön käyttäjän tallennettavaksi.





## KUVIEN LUOKITTELU AINEISTOSTA

Sanoma- ja aikakauslehdet sisältävät tekstin lisäksi myös kuvia, mutta kuvien löytäminen lehtiarkiston tietokannasta on ollut toistaiseksi melko hankalaa. Aiemmin toteutimme yksinkertaisen kuvahaun, jolla pystyi hakemaan esiin sivuja, joissa on jokin kuva. Jos hakuun lisäsi tekstikenttään hakusanan kuvan sisällöstä ja esimerkiksi kuvan kuvateksti sisälsi sanan, haku löysi halutunlaisia kuvia. Toisaalta pelkkä tekstihaku osuu myös kaikkeen kuvan kanssa samalla sivulla olevaan tekstiin, joten haussa tuli helposti myös vain tekstiä, jossa hakusana oli, ilman että sivulla olisi ollut hakusanaan liittyvää kuvaa.

Hankkeen aikana tutkimme mahdollisuuksia luokitella kuvia koneoppimisella. Päädyimme käyttämään melko yksinkertaista kuvien luokittelua, jossa kuvat on jaoteltu tekstimainoksiin, piirroksiin/painokuviiin, nuotteihin, maalauksiin, valokuviiin, karttoihin ja roskeihin. Opetimme kuvitusten luokittukseen TensorFlow-ohjelmiston<sup>4</sup>. Kun sovelsimme opetettua luokittinta

<sup>4</sup> <https://www.tensorflow.org/>

uuteen noin 50 000 kuvan aineistoon, saimme luokittimen arvioiksi tarkkuudeksi noin 70-75 % käymällä lävitse manuaalisesti noin 10 000 luokittimen luokittellemaa kuvaa. Kuvassa 5 on esimerkki luokittimen antamista luokituksista koeaineistolle. Ylimpänä luokituslistassa on korkeimman pisteetyksen saanut arvio, joka valittaisiin kuvan luokitukseksi. Neljästä esimerkin kuvasta kolme on saanut oikean luokituksen.

																																																																																																																																			
kuvat/kansio116/ik00040_1906-12_12_ai-0001_pic01.jpg	kuvat/kansio116/ik00040_1904-12_12_ai-00008_pic01.jpg	kuvat/kansio116/ik00038_1910-03-15_e_019_pic01.jpg	kuvat/kansio116/ik00038_1909-02-01_3_007_pic01.jpg																																																																																																																																
<table border="1"> <thead> <tr> <th>Luokitus</th> <th>Score</th> <th>Malli</th> <th>Korjattu luokitus</th> </tr> </thead> <tbody> <tr> <td>piirros painokuva</td> <td>0.51501</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>nuotit</td> <td>0.30018</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>roskat</td> <td>0.09587</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td><b>kartat</b></td> <td><b>0.04747</b></td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>valokuva</td> <td>0.01937</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>teksti mainos</td> <td>0.01156</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>maalaus</td> <td>0.01055</td> <td>classif2</td> <td>ei korjausta</td> </tr> </tbody> </table>	Luokitus	Score	Malli	Korjattu luokitus	piirros painokuva	0.51501	classif2	ei korjausta	nuotit	0.30018	classif2	ei korjausta	roskat	0.09587	classif2	ei korjausta	<b>kartat</b>	<b>0.04747</b>	classif2	ei korjausta	valokuva	0.01937	classif2	ei korjausta	teksti mainos	0.01156	classif2	ei korjausta	maalaus	0.01055	classif2	ei korjausta	<table border="1"> <thead> <tr> <th>Luokitus</th> <th>Score</th> <th>Malli</th> <th>Korjattu luokitus</th> </tr> </thead> <tbody> <tr> <td><b>kartat</b></td> <td><b>0.67425</b></td> <td>classif2</td> <td>piirros painokuva</td> </tr> <tr> <td>piirros painokuva</td> <td>0.13584</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>nuotit</td> <td>0.10688</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>teksti mainos</td> <td>0.03239</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>roskat</td> <td>0.02225</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>valokuva</td> <td>0.01857</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>maalaus</td> <td>0.00982</td> <td>classif2</td> <td>ei korjausta</td> </tr> </tbody> </table>	Luokitus	Score	Malli	Korjattu luokitus	<b>kartat</b>	<b>0.67425</b>	classif2	piirros painokuva	piirros painokuva	0.13584	classif2	ei korjausta	nuotit	0.10688	classif2	ei korjausta	teksti mainos	0.03239	classif2	ei korjausta	roskat	0.02225	classif2	ei korjausta	valokuva	0.01857	classif2	ei korjausta	maalaus	0.00982	classif2	ei korjausta	<table border="1"> <thead> <tr> <th>Luokitus</th> <th>Score</th> <th>Malli</th> <th>Korjattu luokitus</th> </tr> </thead> <tbody> <tr> <td>valokuva</td> <td>0.84317</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>maalaus</td> <td>0.06499</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>piirros painokuva</td> <td>0.05749</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>roskat</td> <td>0.01985</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>teksti mainos</td> <td>0.00662</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td><b>kartat</b></td> <td><b>0.00463</b></td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>nuotit</td> <td>0.00326</td> <td>classif2</td> <td>ei korjausta</td> </tr> </tbody> </table>	Luokitus	Score	Malli	Korjattu luokitus	valokuva	0.84317	classif2	ei korjausta	maalaus	0.06499	classif2	ei korjausta	piirros painokuva	0.05749	classif2	ei korjausta	roskat	0.01985	classif2	ei korjausta	teksti mainos	0.00662	classif2	ei korjausta	<b>kartat</b>	<b>0.00463</b>	classif2	ei korjausta	nuotit	0.00326	classif2	ei korjausta	<table border="1"> <thead> <tr> <th>Luokitus</th> <th>Score</th> <th>Malli</th> <th>Korjattu luokitus</th> </tr> </thead> <tbody> <tr> <td>valokuva</td> <td>0.94623</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>maalaus</td> <td>0.03275</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>roskat</td> <td>0.00843</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>piirros painokuva</td> <td>0.00721</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>teksti mainos</td> <td>0.00269</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td>nuotit</td> <td>0.00142</td> <td>classif2</td> <td>ei korjausta</td> </tr> <tr> <td><b>kartat</b></td> <td><b>0.00128</b></td> <td>classif2</td> <td>ei korjausta</td> </tr> </tbody> </table>	Luokitus	Score	Malli	Korjattu luokitus	valokuva	0.94623	classif2	ei korjausta	maalaus	0.03275	classif2	ei korjausta	roskat	0.00843	classif2	ei korjausta	piirros painokuva	0.00721	classif2	ei korjausta	teksti mainos	0.00269	classif2	ei korjausta	nuotit	0.00142	classif2	ei korjausta	<b>kartat</b>	<b>0.00128</b>	classif2	ei korjausta
Luokitus	Score	Malli	Korjattu luokitus																																																																																																																																
piirros painokuva	0.51501	classif2	ei korjausta																																																																																																																																
nuotit	0.30018	classif2	ei korjausta																																																																																																																																
roskat	0.09587	classif2	ei korjausta																																																																																																																																
<b>kartat</b>	<b>0.04747</b>	classif2	ei korjausta																																																																																																																																
valokuva	0.01937	classif2	ei korjausta																																																																																																																																
teksti mainos	0.01156	classif2	ei korjausta																																																																																																																																
maalaus	0.01055	classif2	ei korjausta																																																																																																																																
Luokitus	Score	Malli	Korjattu luokitus																																																																																																																																
<b>kartat</b>	<b>0.67425</b>	classif2	piirros painokuva																																																																																																																																
piirros painokuva	0.13584	classif2	ei korjausta																																																																																																																																
nuotit	0.10688	classif2	ei korjausta																																																																																																																																
teksti mainos	0.03239	classif2	ei korjausta																																																																																																																																
roskat	0.02225	classif2	ei korjausta																																																																																																																																
valokuva	0.01857	classif2	ei korjausta																																																																																																																																
maalaus	0.00982	classif2	ei korjausta																																																																																																																																
Luokitus	Score	Malli	Korjattu luokitus																																																																																																																																
valokuva	0.84317	classif2	ei korjausta																																																																																																																																
maalaus	0.06499	classif2	ei korjausta																																																																																																																																
piirros painokuva	0.05749	classif2	ei korjausta																																																																																																																																
roskat	0.01985	classif2	ei korjausta																																																																																																																																
teksti mainos	0.00662	classif2	ei korjausta																																																																																																																																
<b>kartat</b>	<b>0.00463</b>	classif2	ei korjausta																																																																																																																																
nuotit	0.00326	classif2	ei korjausta																																																																																																																																
Luokitus	Score	Malli	Korjattu luokitus																																																																																																																																
valokuva	0.94623	classif2	ei korjausta																																																																																																																																
maalaus	0.03275	classif2	ei korjausta																																																																																																																																
roskat	0.00843	classif2	ei korjausta																																																																																																																																
piirros painokuva	0.00721	classif2	ei korjausta																																																																																																																																
teksti mainos	0.00269	classif2	ei korjausta																																																																																																																																
nuotit	0.00142	classif2	ei korjausta																																																																																																																																
<b>kartat</b>	<b>0.00128</b>	classif2	ei korjausta																																																																																																																																

Kuva 5. Kuvien luokitusta

## DIGI.KANSALLISKIRJASTO.FI -KÄYTTÖLIITTYMÄÄN TEHDYT UUDISTUKSET

Nimienpoiminnan ja tekstintunnistuksen validointia varten Digiin on lisätty uusi ALTO XML -standardin skeema (ALTO Editorial Board, 2013/2019)<sup>5</sup>. Standardin päivittyessä Kansalliskirjaston Digin yhteyteen sijoitettu skeema mahdollistaa helpon validoinnin ja tarvittaessa jopa lisäräätelöntien jatkokehityksen.

Olemme luoneet myös uuden työkaluohjelman, jolla rikastetut XML-tiedostot (REOCR, uusi optinen luku ja NER, nimien tunnistus) voidaan tuoda Digiin käytettäväksi sekä haussa että käyttöliittymässä. Ohjelma on lisätoiminto aineistojen tuomisen työkaluohjelmaan. Ohjelma mahdollistaa korvaavien ALTO XML -tiedostojen antamisen yhden hakemiston kautta, josta aineistot sijoitetaan alkuperäisten XML-tiedostojen korvaajiksi. Prosessi on toistettavissa ja saman tiedoston voi korvata useitakin kertoja, jos näemme tarpeelliseksi rikastaa XML-tiedostoja lisää myöhemmin. Korvattu XML-tiedosto korvaa hakuindeksissä nykyisin olevan version, mikä parantaa tekstihakujen osuvuutta.

<sup>5</sup> [https://digi.kansalliskirjasto.fi/schemas/alto/alto-4-0-nlf\\_reocr.xsd](https://digi.kansalliskirjasto.fi/schemas/alto/alto-4-0-nlf_reocr.xsd)

Nimien poiminta visualisoidaan Digin sisältösivulla siten, että osuneet paikannimet ja henkilönnimet ovat helpommin eroteltavissa sivun tekstistä. Nimet ympäröidään erivärisellä kehyksellä. Oletuksena nimet eivät ole korostettuna, vaan käyttäjä voi valita toiminnon halutessaan. Samoin artikkelipoiminnan lopputuloksena Digiin merkityt leikkeet on muutettu niin, että sisältösivulla kaikki leikkeet eivät ole oletuksena enää näkyvissä, vaan erillisen toiminnon takana, josta artikkelit saa halutessaan näkyviin. Leikkeiden avainsanalistauksissa on siirrytty käyttämään yleisen suomalaisen ontologian (YSO)<sup>6</sup> sanastoa ja paikannimilistää on merkittävästi kasvatettu.

Artikkelipoimintaan on kehitetty ohjelma, jolla PIVAJ'n antamat artikkelien alueet voidaan yhdistää ja siten viedä yhtenäisempinä leikealueina Digiin. Artikkelit näkyvät Digissä leikemerkintöinä. Digin leikemerkintäsivulle on tehty uusi toiminto, jolla yksittäisen leikkeen teksti tunnistetaan hyödyntäen Digissä olevaa sivun tekstintunnistusta. Siinä kohdistetaan leikkeen koordinaatit ALTO XML -tiedostossa oleviin koordinaatteihin, ja siten yhdistetyistä alueista saadaan koko leikealueen teksti. Leikealueen tekstintunnistuksen oikeellisuus riippuu edelleen alkuperäisen tekstintunnistuksen laadusta ja siitä, kuinka tarkasti eri palstat on leikealueeseen maalattu.

## TUTKIMUSYHTEISTYÖ

Kansalliskirjaston osaprojekti on tehnyt tutkimusyhteistyötä monien tahojen kanssa. Merkittävintä yhteistyö on ollut Suomen Akatemian rahoittaman Turun ja Helsingin yliopiston historian tutkijoiden sekä tietojenkäsittelijöiden COMHIS-projektin kanssa, jossa Kansalliskirjasto on myös mukana. COMHIS on muun muassa tuottanut Kansalliskirjaston lehtiaineistoista tietokannan, josta voidaan etsiä lehtitekstien toistoja<sup>7</sup>, ja tutkinut toistojen tyyppiä.

Kieliteknologian soveltamisessa on tehty yhteistyötä FIN-CLARIN -konsortion kanssa lähinnä optisen luvun parantamisessa. PIVAJ-ohjelmiston käyttöön ottamisessa ja käytössä teimme yhteistyötä Rouenin yliopiston LITIS-laboratorion kanssa. Löydettyjen paikannimien linkityskokeilussa tehtiin yhteistyötä Aalto-yliopiston tutkijoiden kanssa. Kuvien luokituksessa on tehty yhteistyötä CSC – Tieteen Tietotekniikan Keskus Oy:n tutkijoiden kanssa.

---

<sup>6</sup> <https://finto.fi/ysol/fi/>

<sup>7</sup> <http://comhis.fi/clusters>

## LÄHTEET

ALTO Editorial Board. 2019. Documentation and use cases for ALTO XML. Saatavissa: <https://github.com/altotml/documentation> [viitattu 29.4.2019].

Clausner, C., Pletshacher, S. & Antonacopoulos, A. 2011. Scenario Driven In-Depth Performance Evaluation of Document Layout Analysis Methods. 2011 International Conference on Document Analysis and Recognition (ICDAR). Saatavissa: DOI: 10.1109/ICDAR.2011.282 [viitattu 16.4.2019].

Dunning, A. 2012. European Newspaper Survey Report. PDF-dokumentti. Saatavissa: <http://www.europeana-newspapers.eu/wp-content/uploads/2012/04/D4.1-Europeana-newspapers-survey-report.pdf> [viitattu 12.4.2019].

Hebert, D., Palfray, T., Nicolas, T., Tranouez, P. & Paquet, T. 2014. Automatic article extraction in old Newspapers Digitized Collections. DATECH '14 Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage, 3–8. Saatavissa: <http://dl.acm.org/citation.cfm?id=2595195> [viitattu 12.4.2019].

Holley, R. 2009. How good can it get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs. D-Lib Magazine, 15(3/4). Saatavissa: <http://www.dlib.org/dlib/march09/holley/03holley.html> [viitattu 12.4.2019].

Kettunen, K., Mäkelä, E., Kuokkala, J., Ruokolainen, T. & Niemi, J. 2016. Modern Tools for Old Content – in Search of Named Entities in a Finnish OCRed Historical Newspaper Collection 1771–1910. Teoksessa Krestel, R., Mottin, D. & Müller, E. (toim.) LWDA 2016, Lernen, Wissen, Daten, Analysen 2016. Saatavissa: <http://ceur-ws.org/Vol-1670/> [viitattu 12.4.2019].

Kettunen, K. & Ruokolainen, T. 2017. Names, Right or Wrong: Named Entities in an OCRed Historical Finnish Newspaper Collection. DATECH 2017. Saatavissa: <https://dl.acm.org/citation.cfm?id=3078084> [viitattu 16.4.2019].

Kettunen, K., Ruokolainen, T., Liukkonen, E., Tranouez, P., Antelme, D. & Paquet, T. 2019. Detecting Articles in a Digitized Finnish Historical Newspaper Collection 1771–1929: Early Results Using the PIVAJ Software. DATECH 2019. [viitattu 16.4.2019].

Kettunen, K. & Koistinen, M. 2019. Open Source Tesseract in Re-OCR of Finnish Fraktur from 19th and Early 20th Century Newspapers and Journals – Collected Notes on Quality Improvement. DHN 2019. Saatavissa: [http://ceur-ws.org/Vol-2364/25\\_paper.pdf](http://ceur-ws.org/Vol-2364/25_paper.pdf) [viitattu 16.4.2019].

La Mela, M., Tamper, M. & Kettunen, K. 2019. Finding Nineteenth-century Berry Spots: Recognizing and Linking Place Names in a Historical Newspaper Berry-picking Corpus. DHN2019, CEUR Workshop Proceedings. Saatavissa: [http://ceur-ws.org/Vol-2364/27\\_paper.pdf](http://ceur-ws.org/Vol-2364/27_paper.pdf) [viitattu 16.4.2019].

Neudecker, C., Wilms, L., Faber, W. J. & van Veen, T. 2014. Large-scale Refinement of Digital Historic Newspapers with Named Entity Recognition. IFLA 2014. Saatavissa: [http://www.ifla.org/files/assets/newspapers/Geneva\\_2014/s6-neudecker\\_faber\\_wilms-en.pdf](http://www.ifla.org/files/assets/newspapers/Geneva_2014/s6-neudecker_faber_wilms-en.pdf) [viitattu 12.4.2019].

Piotrowski, M. 2012. Natural Language Processing for Historical Texts. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Pääkkönen, T., Kervinen, J., Nivala, A., Kettunen, K. & Mäkelä, E. 2016. Exporting Finnish Digitized Historical Newspaper Contents for Offline Use. D-Lib Magazine, July/August. Saatavissa: <http://www.dlib.org/dlib/july16/paakkonen/07paakkonen.html> [viitattu 12.4.2019].

Pääkkönen, T., Kervinen, J. & Kettunen, K. 2018. Digitisation and Digital Library Presentation System – A Resource-Conscientious Approach. DHN18. Saatavissa: [https://helda.helsinki.fi/bitstream/handle/10138/299444/P\\_kk\\_nen\\_Digitisation\\_and\\_Digital\\_Library\\_Presentation\\_System\\_A\\_Resource\\_Conscientious\\_Approach\\_117\\_a.pdf?sequence=1](https://helda.helsinki.fi/bitstream/handle/10138/299444/P_kk_nen_Digitisation_and_Digital_Library_Presentation_System_A_Resource_Conscientious_Approach_117_a.pdf?sequence=1) [viitattu 16.4.2019].

Ruokolainen, T. & Kettunen, K. 2018. À la recherche du nom perdu – Searching for Named Entities with Stanford NER in a Finnish Historical Newspaper and Journal Collection. DAS 2018. Saatavilla: [https://das2018.cvl.tuwien.ac.at/media/filer\\_public/85/fd/85fd4698-040f-45f4-8fcc-56d66533b82d/das2018\\_short\\_papers.pdf](https://das2018.cvl.tuwien.ac.at/media/filer_public/85/fd/85fd4698-040f-45f4-8fcc-56d66533b82d/das2018_short_papers.pdf) [viitattu 16.4.2019].

The “State of the Art”: A Comparative Analysis of Newspaper Digitization to Date. 2015. PDF-dokumentti. Saatavissa: [http://www.crl.edu/sites/default/files/d6/attachments/events/ICON\\_Report-State\\_of\\_Digitization\\_final.pdf](http://www.crl.edu/sites/default/files/d6/attachments/events/ICON_Report-State_of_Digitization_final.pdf) [viitattu 12.4.2019].

# AUTOMATISOINNILLA ETUJA ASIAKIRJAHALLINTAAN

**Anssi Jääskeläinen, TKT, tki-asiantuntija & Tuomo Räisänen, FT, it-asiantuntija,  
Kaakkois-Suomen ammattikorkeakoulu**

Maailmassa tuotettavan tiedon määrä kasvaa eksponentiaalisesti. Digitalia onkin kehittänyt dokumenttien automaattiseen pilkkontaan, analysointiin ja metatiedottamiseen liittyvää avoimeen lähdekoodiin perustuvaa tekniikkaa jo pidempään, mahdollistaen näin esimerkiksi seulonnan osittaisen automatisoinnin. Tässä artikkelissa kuvataan kehitettyä tekniikkaa yleisellä tasolla. Pääasiallisena yhteistyökumppaninamme on ollut Kansallisarkisto, jolta olemme saaneet mm. materiaalia käyttöömme (Kansallisarkisto 2018a). Työn loppupuolella mukaan kehitykseen tuli myös Kirkkohallitus, jonka aineistoihin jo toteutettuja ratkaisujamme testattiin hyvin tuloksin.

## AINEISTO

Kansallisarkistolta saamamme pilottimateriaalit ovat seulonta-asiakirjoja vuosilta 1952–2006. Seulonta-asiakirjat koostuvat julkisen hallinnon toimijan seulontaesitysasiakirjoista (esim. arkistonmuodostussuunnitelma), päätösvalmistelijan laatimasta muistiosta ja seulontapäätösasiakirjasta. Testisettimme koostui noin 75000 JPG-kuvatiedostosta (7 hm A4-koon paperiasiakirjoja). (Kansallisarkisto 2018a; 2018b)

## OCR

JPG-kuvatiedostoissa ei ollut mitään koneluettavaa tekstisisältötietoa valmiina, vaan ensimmäinen vaihe työssä oli kuvatiedostojen OCR-luku. Ratkaisuksi toteutimme Python-ohjelman, joka lukee kansiot läpi ja kohdatessaan kuvatiedoston siirtää sen moniprosessointimoduulille jatkokäsittelyyn. Jos skripti puolestaan havaitsee PDF-tiedoston, jossa ei ole OCR-tietoa, se muutetaan ensin OCR-lukua varten sivukohtaisesti PNG-kuvatiedostoiksi, joita käyttämämme avoimen lähdekoodin Tesseract 4 osaa käsitellä. Jokainen kohdattu kuvatiedosto OCR-luetaan ja lopuksi samaan kokonaisuuteen kuuluvat OCR-tiedon sisältämät tiedostot yhdistetään yhdeksi OCR-tiedon sisältäväksi PDF-tiedostoksi. Esimerkiksi 46032 JPG-tiedoston, joista 30765 sisälsi Tesseractin mielestä tekstiksi luokiteltavaa aineistoa, luenta kesti runsaat 23 tuntia, kun lukuprosessiin käytettiin 32:ta virtuaalista CPU:ta. Tästä saadaan keskimääräiseksi lukuajaksi noin 2,7s per tekstisivu ja 1.8s per sivu, jos mukaan lasketaan myös tyhjä. Tesseract ei poista tyhjiä sivuja automaattisesti, joten kehitimme vielä erillisen Python-skriptin, joka pyrkii poistamaan nämä tyhjä sivut. Tämän skriptin ajo kesti noin 15 minuuttia ja jäljelle jäi 22817 sivua, jotka sisälsivät tekstiä. Suuri pudotus määrässä johtuu siitä, että Kansallisarkistolla skannatuissa asiakirjoissa on myös kääntöpuoli mukana ja tässä

materiaalissa kääntöpuoli oli aina tyhjä. Loput erosta selittyi sillä, että arkistokansioissa on aina kansilehtiä ja erinäisiä tyhjiä sivuja mukana.

## PILKONTA

Saamamme materiaali oli skannattu arkistomappi kerrallaan yhdeksi kokonaisuudeksi, joten varsinainen tehtävämme oli erotella toisistaan asiakokonaisuudet sekä erotella jokaisesta asiakokonaisuudesta vielä esimerkiksi Kansallisarkiston sisäiset muistiot. Vaikka tehtävä voi kuulostaa helpolta, se on kuitenkin kaikkea muuta. Seuraavassa muutamia syitä:

- Tunnistesivun, esityslistan, päätöksen, muistion ja liitteiden järjestys ei ole vakio
- Jokaista edellä mainitun listan edustajia voi olla nolla, yksi tai useampia
- Sivujen ulkoasu ja formaatti ovat muuttuneet monesti vuosien saatossa
- Eri osien aluissa ja loppuissa ei ole yksiselitteisiä tunnisteita

Lähdimme kuitenkin yrittämään ajatuksenamme se, että koodausta lukuun ottamatta mitään ei tehtäisi käsin.

Kirjaimien tunnistaminen OCR:n avulla perustuu opetettuun tekoälyyn sekä taustalla tapahtuvaan kuvamanipulaatioon. Tämä toimii vaihtelevasti riippuen vahvasti digitoidun aineiston laadusta. Käsin kirjoitettu materiaali oli rajattu tämän projektin ulkopuolelle, mutta odotamme mielenkiinnolla kansainvälisen READ-hankkeen<sup>8</sup> tuloksia ja sen hyödyntämismahdollisuuksia. Tässä kokeilussamme Kansallisarkiston kanssa tunnistaminen tapahtui ainoastaan konekirjoitetulle materiaalille algoritmiperusteisesti.

Kansallisarkiston toiveesta aineisto jaettiin kahteen osaan. OSA1 sisälsi seulontapäätöksiä vuosilta 1952-1980. Käsiteltäviä PDF-tiedostoja oli tässä tapauksessa vain viisi, jotka kuitenkin sisälsivät satoja päätöksiä liitteineen. OSA2 puolestaan koostui myöhemmistä vajaasta tuhannesta tiedostosta, jotka sisälsivät pääasiassa yksittäisten päätösten asiakirjoja. Saimme Kansallisarkistolta tähän aineistoon referenssidatan, joka sisälsi kunkin seulontapäätöksen diaarinumeron. Näin tuloksista voitiin laskea suoraan "osumat". Päätöksissä liitteineen on tyypillisesti seulontapäätösten hakemuksia sisältäen eri toimijoiden diaarinumeroita, joten yllä kuvattu tiedostojen pilkkominen on tehtävä. Jos ei tunnisterä päätöstä, ei luonnollisesti voida löytää diaariakaan. Molemmille osille tehtiin erilliset skriptit.

## TULOKSIA OSA1

Pöytäkirjoja lukiessa tehtiin se havainto, että sanat "valtiorakiston" ja "pöytäkirja" esiintyivät aina päätösten alussa. OCR-luku ei kuitenkaan ole aina tulkinnut sanoja oikein, vaan sana "pöytäkirja" saattaa saada muodon "pötykirja". Sitä mukaa kun näitä OCR-virheitä löydettiin, ne tallennettiin erilliseen listaan, joka tarkastettiin aineistoa läpikäydessä. Tarkoituksena

---

<sup>8</sup> <https://cordis.europa.eu/project/rcn/198756/factsheet/en>

oli myöskin nimetä pöytäkirjat numeron ja päivämäärän mukaan. Pöytäkirjan numeroiden etsinnästä luovuttiin, koska aineistossa ei ollut kovinkaan johdonmukaista käytäntöä niiden merkitsemiseen. Lisäksi osaan aineistoa oli numerointi merkitty jälkikäteen käsin.

Päivämäärän suhteenkin oli vaihtelua. Alkupään aineistossa päivämäärä oli ilmoitettu kirjoittamalla (ei numeroina). Toki tähänkin voisi kehittää oman algoritminsa, mutta aineiston koon ja projektin keston huomioon ottaen tähän ei ryhdytty. Päivämäärää etsittiin siten vain numeromuodossa. Kun aineisto oli ensin pilkottu, etsittiin sieltä päivämäärää. Tämän avulla voitiin sitten muodostaa sisällysluettelo. Numerointi korvattiin pöytäkirjan järjestyksellä tiedostossa ja jos päivämäärä löytyi, sekin tieto lisättiin sekä sisällysluetteloon että pöytäkirjan nimeen. Pilkonta onnistui kutakuinkin täydellisesti, mutta nimeämiseen tarvittavassa päivämäärässä jouduttiin tyytymään noin 75 %:in päivämääräistä.

## TULOKSIA OSA2

Kuten kappaleen alussa kuvailtiin, tämä aineisto oli huomattavasti mutkikkaampi. Päätös saattoi "luurata" missä millä tahansa sivuvälillä useamman sadan sivun tiedostossa. Joissakin tapauksissa apuna olivat sanat "luonnos" tai "päätösluonnos", mutta havaitsimme, että parempi oli yrittää Sherlock Holmes -taktiikkaa "What ever remains, must be the seulontapäätös". Esityslistojen, muistoiden ja joissakin tapauksissa liitteidenkin alkukohdat löytyivät avainsanoilla. Ei kuitenkaan löytynyt mitään tapaa päätellä automaattisesti, milloin esityslista tai muistio loppuu, puhumattakaan liitteistä.

Yllättävää kyllä, päätöksen loppuosassa oli aina kahden henkilön allekirjoitus, jota edelsi heidän tittelinsä. Valitettavasti joskus näin oli myös viranomaisen toimittamassa seulontaesityksessä. Muistioita ja liitteitä saattoi olla useita sikiä sokin ympäri tiedostoa.

Jotakin ensimmäisen osan kokemuksista opittiin: siellä missä jokin osa alkaa, edeltävä osa loppuu. Tässä tapauksessa päädyimme ensin luomaan aineistosta sisällysluettelon, jonka avulla sitten varsinainen pilkonta onnistuu. Referenssin kannalta tärkeää diaarinumeroa etsittiin sitten päätöksen ensimmäiseltä sivulta. Joskus tosin diaarinumero saattoi olla aineistossa ihan missä vain, mutta näiden tapausten erittelemine jätettiin TODO-listalle. Diaarinumero talletettiin erilliseen tiedostoon ja 95 % tarkkuudella toteutettu algoritmi jonkinlaisen diaarinumeron löysikin, ja tätä vertailtiin referenssiin. Muutamien kokeilujen kautta onnistuttiin saavuttamaan seuraavanlaiset tulokset. Kun algoritmin antamia diaarinumeroita verrattiin referenssiaineistoon, havaittiin, että diaarinumerot täsmäävät noin 63 % tarkkuudella.

Entä mistä virheet johtuivat? Ensihavaintona oli se, että useissa tapauksissa referenssimerkintä ei vastannut seulontapäätöksessä olevaa, vaan tunnuksia (esim. KA) oli saatettu lisätä diaariin jälkikäteen tai diaarinumero oli merkattu aineistoon käsin kirjoittamalla. Selvästikään näitä materiaaleja ei ollut tarkoitettu koneellisesti luettaviksi ja vertailtaviksi. Tällaisen ongelman huomioiminen ohjelmoimalla on vähintäänkin vaikeaa ja joskus jopa mahdotonta,



joten suoritimme kokeen, jossa verrattiin vain diaarinumeroita toisiinsa. Tällä muutoksella tunnistamisprosentti parani yli 70 %:n. Seuraava havainto oli se, että joskus numerointikäytäntökin vaihteli. Vuosiluku 2006 saatettiin lyhentää muotoon 06 diaarinumerossa tai sitten päinvastoin. Näitä tapauksia havaittiin referenssiaineistosta muutamia kymmeniä. Olisi toki mahdollista toteuttaa skripti, joka yhtenäistää vuosilukukäytännön, mutta tässä projektissa ei sellaista toteutettu. Jos tätä vuosilukuongelmaa ei oteta huomioon, tulokset “paranevat” yli 80 %:n tarkkuuteen. Loput virheistä johtuivat siitä, että algoritmi ei tunnistanut päätöstä tai diaaria. Nämä virheet voivat johtua OCR-virheistä, algoritmin logiikasta tai ohjelmointivirheistä.

Seuraavaksi selvitimme, miten tarkkuus vaihtelee aineiston iästä riippuen. Havaitimme, että virheitä syntyi eniten 80-luvun aineistosta. Syynä on kopioidun aineiston laatu, joka on parantunut ajan myötä.

*Taulukko 1. Diaarinumeroiden löytymistarkkuus vuosikymmenittäin*

	80-luku	90-luku	00-
Lkm	300	300	156
Osumat	78%	83%	95%

Aineisto oli digitoitu käyttäen KODAK- ja IBML-laitteistoja. Emme havainneet laitteistoriippuvaisia eroja omissa tuloksissamme.

Mitä suorituskykyyn tulee, yhden digitoidun hyllymetrin muuttaminen ensin PDF:ksi ja sen analysointi kesti käytettävissä olevilla laitteistoilla useita tunteja.

## JOHTOPÄÄTÖKSET

Osa1:n tulokset osoittavat, että koneellisella pilkkonnalla voidaan saavuttaa huomattavia kustannussäästöjä. Pöytäkirjojen erottelu käsin isosta tiedostosta on aikaa vievää puuhaa. Osa2:n tulokset taas mahdollistavat seulontapäätösten hakemisen digitoidusta aineistosta, siltä osin kuin oikea diaarinumero löytyy. Nykykäytäntöön verrattuna edut ovat ilmeiset. Kehittämiämme algoritmeja voidaan käyttää sekä aineiston fyysiseen pilkkontaan että aineiston ns. metatiedon keräämiseen, jolloin varsinainen tiedonhaku voidaan toteuttaa teknisesti alkuperäisestä tiedostosta.

## SISÄLTÖANALYYSI

Kun kokonaisuus on saatu pilkkottua halutunlaiseksi ja olisi aika tehdä asiasanoittaminen, halukkaita tuskin olisi jonoksi asti. Seuraava perusongelmamme on siis automaattinen asiasanoittaminen dokumenttien tai asiakirjojen sisällön perusteella ilman, että kenenkään tarvitsee itse lukea niitä itse läpi.



Analysaattorimme toimii tällä hetkellä pdf-tiedostojen kanssa ja tunnistaa siitä kuvassa 1 esitettyjen asioiden lisäksi myös sähköpostiosoitteet, URL-osoitteet, suomalaiset yritystunnukset, puhelinnumerot sekä luottokorttinumerot. Nämä eroteltaisiin raportissa omana taulukkonaan.

Tätä ratkaisuumme on mahdollista testata joko demosivustollamme osoitteessa digitalia.xamk.fi tai uuden APIimme kautta, josta voi kysellä lisää Digitalian Xamk-tiimiltä.

## ANONYMISOINTI

Lainsäädäntö rajoittaa tiedon jakamista julkisessa Internetissä. Tilanne helpottuu huomattavasti, kun henkilötiedoiksi luokiteltavat asiat poistetaan aineistosta.

Rakensimme avoimen lähdekoodin ohjelmistojen ja Pythonin yhdistelmällä anonymisointityökalun, jonka toiminta perustuu NER-tunnistukseen, avoindata.fi -palvelussa saatavilla oleviin etunimi- ja sukunimilistoihin sekä käyttäjän itsensä määrittämiin sanalistoihin. Näiden yhdistelmällä päästään anonymisoinnissa suhteellisen hyviin tuloksiin, mutta lopullinen tarkastaminen on kuitenkin syytä jättää edelleen ihmiselle.

### References

- [1] ██████████ ██████████. Personal Archiving: Preserving Our Digital Heritage (Medford, NJ, 2013)
- [2] L. McCay-Peet, ██████████ Quan-Haase, What is social media and what questions can social media research help us answer? In L. ██████████ & ██████████ Quan-Haase (Eds), ██████████ SAGE Handbook of Social Media Research Methods, pp. 13-26. (SAGE, LA, 2017)
- [3] ██████████ ██████████, L. ██████████. OCR: Unleash the hidden information, Proc. Archiving 2018, pg. 83. (2018).
- [4] ██████████ ██████████, I. ██████████ ██████████, J.D. ██████████. Data Storage, Curation and Preservation. In L. ██████████ & ██████████ Quan-Haase (Eds), ██████████ SAGE Handbook of Social Media Research Methods, pp. 161-176. (SAGE, LA, 2017)

### Author Biography

██████████ ██████████ has an IT MSc. (2005) and a PhD (2011) from Lappeenranta University of Technology. ██████████ has an extensive knowledge of user experience and usability. His current interests are in information governance, format migration and ██████████ development.

██████████ ██████████ holds a PhD (2008) from Lappeenranta University of Technology. She is an experienced researcher and trainer in knowledge management, open communities, social technologies, and social media. Her current interests are in the field of digital communication research and the related methods.

### Kuva 2. Esimerkki anonymisoinnista

Kuvassa 2 on esimerkki anonymisoiduista lähteistä ja kirjoittajien biographyistä. Punaisten laatikoiden alla olevat tekstit on tunnistettu NER-työkalulla, mustien laatikoiden alla olevat etu- ja sukunimilistoilla ja sinisten laatikoiden alla oleva sana puolestaan käyttäjän itsensä määrittämän sanalistan avulla. Kuten jo tästä esimerkistä voi havaita, kaikki nimet eivät

tunnistu oikein ja esimerkiksi 'He' sana ylimmässä biographyssä on tunnistunut joko etu- tai sukunimilistan perusteella nimeksi. Kehitettävää tällä saralla siis löytyy vielä, mutta jo tällainenkin tarkkuus helpottaa huomattavasti käsityönä tehtävää anonymisointia.

Tällä hetkellä tekninen ratkaisumme perustuu anonymisoitavien kohteiden koordinaattien avulla tapahtuvaan laatikolla peittämiseen. Tämä ei kuitenkaan riitä poistamaan laatikon alla olevaa tekstiä, vaan seuraavassa vaiheessa tallennamme koko sivun kuvaksi ja OCR-luemme sen uudelleen. Vaikka testeissämme emme ole havainneet OCR-laadun heikentymistä, ratkaisu on kuitenkin hieman kömpelö ja tutkimmekin parhaillaan mahdollisuutta poistaa sana suoraan PDF-tiedoston sisältä rikkomatta muuta PDF:n rakennetta. Tätä varten kokeilimme open source -pakettia nimeltä pdf-redactor. PDF-tiedostojen sisäinen rakenne on todella monimutkainen ja paketissa on lukuisia rajoitteita. Se osaa kuitenkin korvata merkkijonon toisella mahdollistaen anonymisoinnin. Alustavien testien perusteella tämä ratkaisu on käytökelpoinen. Etuna tässä lähestymistavassa on aineiston käsittelyn nopeutuminen.

## **SÄHKÖPOSTIKONVERSIO**

Tästä asiasta olemme raportoineet aiemminkin useassa eri kanavassa, mm. Jääskeläinen ja Kokkonen (2017). Siksi tässä kappaleessa on vain lyhyt nykytilan päivitys. Sähköpostikonvertteriamme voi testata osoitteessa [digitalia.xamk.fi](https://digitalia.xamk.fi) sekä uuden APIimme kautta, josta voi kysellä lisätietoja Digitalian Xamk-tiimiltä. Yhteistyössä Päivälehdien arkiston ja Disec Oy:n kanssa olemme korjailleet konvertterista virheitä ja tehostaneet sen toimintaa.

## LÄHTEET

Jääskeläinen, A. & Kokkonen, T. 2017. Sähköpostien ihanuus ja kurjuus. Teoksessa Kosonen, M. (toim.), Digitaalinen tieto haltuun: Ratkaisuja digitaalisten aineistojen hallintaan ja käyttöön. Mikkeli: Xamk kehittää 11, Kaakkois-Suomen ammattikorkeakoulu, 48-57.

Kansallisarkisto 2018a. Kansallisarkiston analogisten seulonta-asiakirjojen digitointi- ja tutkimussuunnitelma. Kansallisarkisto, KA/17578/07.01.02.04.00/2018.

Kansallisarkisto 2018b. Massadigitoinnin suunnitteluprojektin loppuraportti. Kansallisarkisto.

# 3D-PELIMOOTTORIN HYÖDYNTÄMINEN HISTORIALLISTEN TAPAHTUMIEN VISUALISOINNISSA

Anssi Jääskeläinen, TKT, tki-asiantuntija, Kaakkois-Suomen ammattikorkeakoulu

Onko pelkän tekstin lukeminen ja siitä asioiden ymmärtäminen haastavaa, tylsää tai muuten vain puuduttavaa? Selviäisikö esimerkiksi paikkakuntakohtainen äänestysaktiivisuus eduskuntavaaleissa helpommin visuaalisesta esityksestä kuin monesta sivusta numerotekstiä? Uskomme, että vastaus on useimmiten kyllä ja sama pätee myös historiallisiin faktoihin.

Jos esimerkin vuoksi kerron, että Helsingin suurpommitusten ensimmäinen hyökkäys tapahtui 6.-7. helmikuuta 1944 kello 18.51-04.57, jolloin pudotettiin noin 7000 pommia, joista noin 2500 osui maahan ja loput mereen, ja että maahan osuneista vain 350 osui kohteeseensa, et tämän tekstin loppuun luettuasi todennäköisesti muista paljoakaan kerrotuista lukemista. Tämä on perusongelma, johon kirjoitetun tekstin kanssa usein törmätään. Vaihtoehtoisia oppimistapoja on monia, mutta tämän työn pohjana on ollut visuaalinen oppiminen ja havainnointi.

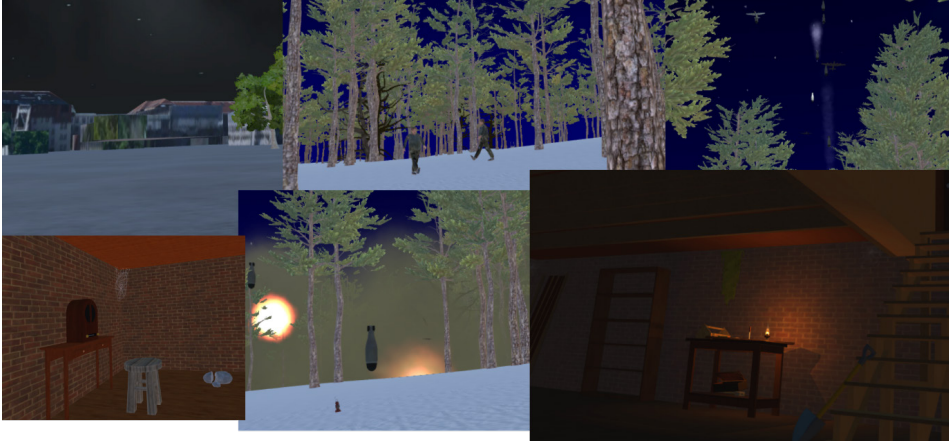
Visuaalista oppimista ja sen mukanaan tuomia hyötyjä on tutkittu paljon. Esimerkiksi McGrath ja Brown (2005) toteavat, että visuaalisten tekniikoiden käyttämisellä on monia etuja tiede- ja insinöörikoulutuksessa, koska ne tukevat tiedon sisäistämistä. Lorenzo ym. (2013) tutkivat, kuinka VR-tekniikan hyödyntäminen auttaa Aspergerin syndroomasta kärsiviä lapsia oppimaan asioita eri tavalla ja selviytymään erilaisista testeistä huomattavasti paremmin kuin ilman VR-tekniikan apua.

Mistään uudesta asiasta ei kuitenkaan ole kysymys. Jo Winn (1993) totesi, että VR-tekniikan avulla opiskelijoille voidaan luoda sellaisia kokemuksia, jotka eivät muuten olisi mahdollisia. Näin VR tukee vahvasti oppimista tuomalla peliin mukaan lisää aisteja. Näiden faktojen pohjalta lähdimme soveltamaan VR-tekniikkaa case-tapaukseemme.

## CASE HELSINGIN SUURPOMMITUKSET

Yksi hankkeen tutkimusteemoista oli olemassa olevan tiedon visualisointi. Päätimmekin testata tätä visualisointia historiallisen aineiston kanssa. Koska kaikenlaisia Excel-käppyröitä ja graafeja on nähty aivan riittävästi, päätimme hyödyntää tällä kertaa pelimoottoritekno-

logiaa sekä VR-laseja. Näiden yhdistelmällä on mahdollista luoda immersiiivinen ja kokijaa aktivoiva 3D-elämys, jossa pelaaja pääsee itse kokemaan ja kuulemaan pommitukset nähdén samalla historiallisia faktoja tapahtumien kulusta. Joitakin ruudunkaappauksia toteutuksesta on esitetty kuvassa 1.



*Kuva 1. Todellisia kuvankaappauksia sovelluksesta*

- Lähtötilanteessa pelaaja pudotetaan aloitusvalikkoon, jossa kerrotaan lyhyesti ohjaus sekä se, mitä itse kokemuksessa tulee tapahtumaan ja että pelaajan pääasiallinen tavoite on selvittää pommituksista hengissä hakeutumalla pommisuojaan ja toivomalla parasta.
- Varsinaisen pelikokemuksen alkaessa pelaaja pudotetaan sattumanvaraiseen paikkaan sattumanvaraiseen pelikenttään, jossa sekä ympäristö, vallitseva valaistus sekä taivas on luotu dynaamisesti. Samanaikaisesti pelaajan pudottamisen kanssa luodaan muutamia pommisuojaan pääsyä merkkeavia kohteita sekä satunnainen määrä erilaisia pommikoneita.
- Pommikoneet päättävät itsenäisesti pommituskohteensa pelialueelta ja jokainen kone kantaa mukanaan tiettyä pommimäärää.
- Pommeilla on satunnainen tuhovoima ja kaikki sen piiriin joutuvat objektit joko särkyvät tai lentävät voiman vaikutuksesta. Tuhovoima vaikuttaa myös pelaajaan, joka kuolee, jos hän ottaa vastaan liian kovan tärskyn.
- Jos pelaaja löytää pommisuojaan merkkeavan objektin, hän siirtyy suoraan pommisuojaan, joka sijaitsee ajalle tyypillisesti talon kellarikerroksessa hieman maan alla. Pommisuojaan on toteutettu sekä vavahdusefekti että katosta putoavan pölyn efekti, jos pommi osuu liian lähelle.
- Pelialueelle luodaan myös satunnainen määrä erinäisiä objekteja, jotka olisivat saataneet esiintyä kyseisen ajankohdan katukuvassa. Lisäksi pelialueelle luodaan NPC (Non Player Character) -hahmoja, jotka vaeltelevat alueella oman tahtonsa ohjaamina. Pommitusten alkaessa nämä hahmot suunnistavat myös pommisuojaan, joten pelaajan

ei tarvitse olla sielläkään yksin. Interaktiota näiden hahmojen kanssa ei toistaiseksi ole kuitenkaan toteutettu.

- Pelisession päätyttyä joko pelaajan kuolemaan tai pommituksen loppumiseen pelaaja näkee faktoja juuri kokemastaan pommituksesta. Mm. kuolonuhrit, pudotetut pommit, käytetyt konetyypit yms. pommituksiin liittyvät tärkeät tiedot tuodaan visuaalisesti pelaajan eteen.

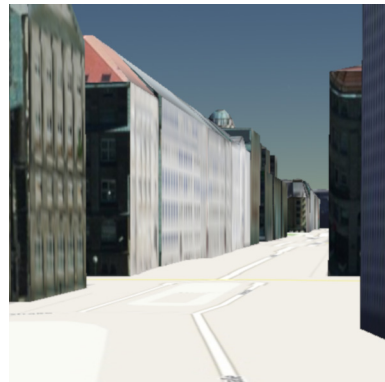
Ilmatorjuntaa tai ilmahälytystä pelialueelle ei vielä kirjoittamishetkellä ole toteutettu, mutta tulevaisuuden suunnitelmiin se on jo kirjattu. Koko 3D-sovellus on lähtökohtaisesti toteutettu dynaamisesti, joten sen laajentaminen uusilla pelialueilla, ympäristöillä ja toiminnoilla on suhteellisen yksinkertaista. Oppimiskokemuksesta voisi tulevaisuudessa tehdä vieläkin kiinnostavamman mahdollistamalla interaktio NPC-hahmojen kanssa. Nämä hahmot voivat kertoa mm. historiallisista faktoista ja siitä, kuinka he ovat kokeneet pommitukset ja sodan sekä näyttää esimerkiksi valokuvia aiempien pommitusten tuhoista. Mahdollisuuksia rajaavat ainoastaan käytössä olevat resurssit.

## TEKNIikka JA TOTEUTUS

Unity-pelimoottorina toteutuksessa käytettiin versiota 2018.3.12f1, myös uusinta 2019 versiota kokeiltiin mutta sen VR-tuki on ilmeisesti vielä hieman puutteellinen ja toteutuksen toimintaan saattamisessa ilmeni ongelmia. VR-tekniikkana käytössä oli Oculus Rift kahdella Oculus Touch -ohjaimella ja kahdella sensorilla. Lisäksi asennettuna Unityyn oli Oculus Integration for Unity:n versio 1.35.

Suurin osa pelialueilla näkyvistä ja esiintyvistä objekteista, kuten lentokoneet, pommit, vankurit, tiesulut, lyhdyt, yms. on kerätty monista vapaasti saatavilla olevista 3D-kirjastoista. Myös pommisuojaissa esiintyvissä objekteissa on hyödynnetty olemassa olevia 3D-malleja. Metsämaaston luomisessa käytetään Unityn C# ohjelmointikieltä. Jokaisella ajokerralla käytetään matemaattisia funktioita sekä satunnaisuutta, jotta metsämaastopelikentästä saadaan aina uniikki. Maastoon sijoitetaan myös puut ja muut objektit satunnaisesti.

Kaupunkimallin pohjana on puolestaan vapaasti saatavilla oleva Berliinin kaupunkimalli<sup>10</sup> (kuva 2), jonka voi ladata Unityn ymmärtämässä DXF- tai 3DS-muodossa. Meidän tapauksessamme ladattu malli tosin pyöräytettiin 3ds Max -ohjelman kautta, jossa mallia kevennettiin ja normalisoitiin ennen

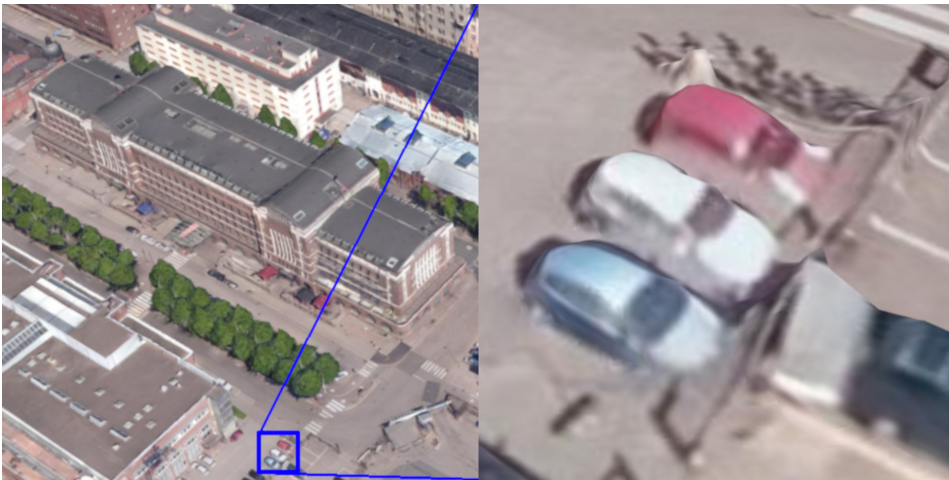


*Kuva 2. Berliinin kaupunkimallia*

<sup>10</sup> <https://www.businesslocationcenter.de/berlin3d-downloadportal>



Unityyn vientiä. Koska kyseessä on Helsingin suurpommitukset, olisi toki ollut luonnollisempaa käyttää Helsingin kaupunkitietomallia, mutta toteutushetkellä saatavilla olevan OBJ-mallin jokainen karttalehti on jaettu 64 pienempään osaan ja Unity-pelimootorissa ei ole keinoa automaattisesti sijoitella näitä karttapalasia oikeille paikoilleen. Fotorealistisen mallin ongelma puolestaan on kehnosta tarkkuudesta. Kuva 3 kertoo hyvin, miksi tätä mallia ei voida hyödyntää käytettäessä First Person Controller -ohjausta, jota meidän toteutuksemme käyttää.



*Kuva 3. Fotorealististen 3D-mallien ongelma*

Kuvasta 3 ilmenee ilmakuvien pohjalta tehtyjen 3D-mallien ongelma, joka tekee niistä käyttökelvottomia toteutuksissa, joissa pelaaja näkee kentän ihmisen tasosta. Kaikki yksityiskohtat puuroutuvat ja tällaiseen pohjaan on käytännössä mahdotonta luoda esim. kunnollista törmäyksen tunnistusta tai navigointiverkkoa NPC-hahmoille. Huomautuksena mainittakoon, että Helsingin sivuille on ilmeisesti lähiaikoina ilmestynyt mahdollisuus ladata alueita samoin kuin Berliinin tapauksessa, joten tulemme piakkoin lisäämään myös Helsingin alueita pommitussimulaattoriimme.

VR-tekniikka vaatii toimiakseen suhteellisen järeän koneen. Vaikka itse toteutukseen olisi riittänyt vaatimattomampikin mylly, suunnittelun, kehittämisen ja testaamisen sujuvoittamiseksi työkoneeksi valittiin kuitenkin tehokas MSI P65 Creator 8RF. Tehokas GeForce® GTX 1070 näytönohjain on täysin VR-kelpoinen nyt ja lähitulevaisuudessakin. Valmiista toteutuksesta tullaan julkaisemaan myös WebGL-versio, jotta ”filistelemaan” pääsee myös ilman VR-kelpoista konetta suoraan selaimenkin kautta. Valmistuessaan WebGL-versio tulee kokeiltavaksi Digitalian demosivuston kautta osoitteessa digitalia.xamk.fi, jossa on jo nyt paljon mielenkiintoista materiaalia liittyen Digitalian toimintaan. Oculus Riftellä toteutusta pääsee halutessaan kokeilemaan esim. Digitalian työhuoneella Xamkin Mikpoli-rakennuksessa tai ennalta sovitussa tapahtumissa, joissa tarvittava laitteisto on mukana.

## LÄHTEET

Lorenzo, G., Pomares, J. & Lledó, A. 2013. Inclusion of immersive virtual learning environments and visual control systems to support the learning of students with Asperger syndrome. *Computers & Education*, 62, pp. 88-101.

McGrath, M. & Brown, J. 2005. Visual Learning for Science and Engineering. *IEEE Computer Graphics and Applications*, 25(5), pp. 56-63.

Winn, W. 1993. A conceptual basis for educational applications of virtual reality. Technical Publication R-93-9, Human Interface Technology Laboratory of the Washington Technology Center. Seattle: University of Washington.

# AVOIMILLA OHJEISTUKSILLA JA OPPAILLA TUETAAN DIGITAALISEN TIEDON PITKÄAIKAISSÄILYTTÄMISTÄ

Teemu Hänninen, it-tradenomi, it-asiantuntija, Kaakkois-Suomen ammattikorkeakoulu

Kansalaisten, pienyhteisöjen, järjestöjen ja sukujen historiallisten aineistojen säilyttäminen on tärkeä tavoite, jonka toteuttamiseksi tarvitaan digiaikana uusia tietoja ja taitoja. Paperiset valokuvat hapertuvat, isovanhempien rakkauskirjeet häviävät jäljettömiin, tärkeät asiakirjat hautautuvat mappeihin, katoavat tai tuhoutuvat. Pelastusoperaation ensimmäinen vaihe on aineistojen digitointi.

Monet tiedostavat jo digitoinnin ja sähköisten arkistojen tärkeyden, mutta tarvittavaa osaamista ei välttämättä löydy. Siksi tarvitaan avoimia oppaita ja ohjeita, jotka valaisevat uusien menetelmien hyötyjä ja käytäntöjä mahdollisimman kansantajuisesti. Lisäksi ohjeiden pitäisi olla helposti löydettävissä ja saatavilla, koska tieto on hyödytöntä, jos se ei löydy tarvitsijaansa.

## DIGITAALISEN PITKÄAIKAISSÄILYTTÄMISEN TÄRKEYS

Emme edelleenkään ole täysin varmoja siitä, kuinka ja miksi Gizan pyramidit rakennettiin. Suurin osa pyramidien rakennusprosessia kuvaavasta tiedosta on kadonnut, joten joudumme päättämään ja arvailemaan syitä ja seurauksia, mikä tekee historiankirjoittamisesta vaikeaa. Kenties egyptiläisillä oli paljon kirjoitettua tietoa, mutta ilman tiedon säilyvyyden turvaamiseksi suoritettuja toimenpiteitä tieto on kadonnut tomuna ja tuhkana historian havinoihin. Kaikkeen eivät egyptiläisetkään siis pystyneet, vaikka monissa asioissa olivatkin sivilisaation edelläkävijöitä. Myös jälkipolvilla olisi pitänyt olla osaamista ja halua turvata arkistoidun tiedon säilyminen ja käytettävyys.

Nykyään elämme aikakautta, jossa ihmisillä on entistä suurempi tarve pystyä säilyttämään tietoa myös digitaalisessa muodossa. Merkittävä osa aineistosta on syntysähköistä. Myös analogisen aineiston muuntaminen digitaaliseksi onnistuu suhteellisen helposti pienelläkin panostuksella ja kevyellä laitteistolla. Vaikka mahdollisuus on olemassa, tarvittavaa osaamista, resursseja ja varsinkaan aikaa ei välttämättä löydy.

Digitaaliset aineistot käyttöön -hankkeessa kehitetyllä yleisohjeistuksella on pyritty kannustamaan aineistojen digitointiin ja yhtenäistämään digitointiprojektin käytäntöjä, mistä on

hyötyä sekä yksityisille kansalaisille että koko arkistoalalle. Ohjeistus on avoimesti saatavilla ja monien erilaisten toimijoiden hyödynnettävissä.

## **DIGISOTKUSTA KOHTI JÄRJESTYSTÄ**

Digitalisaation myötä yksityishenkilöt, yritykset ja yhdistykset ovat alkaneet nähdä sähköisen pitkäaikais säilyttämisen tärkeyden. Hyödyt tiedostetaan ja tahtotila löytyy, mutta toteutus saattaa jäädä uupumaan. Hankkeessa toteutettujen käyttäjähaastattelujen perusteella oli selvää, että varsinkin pienyritykset ja -yhdistykset tarvitsevat opastusta perusasioista lähtien. Digitaalisten aineistojen arkistointi koetaan monimutkaisena sekasotkuna, mitä pahentaa entuudestaan se, jos organisaation tiedonhallinta on ollut epäjärjestelmällistä. Syntysähköisiä asiakirjoja on tallennettu useaan eri paikkaan, esimerkiksi eri pilvipalveluihin, tai ne saattavat olla ainoastaan yhden vastuuhenkilön koneella ja tuhoutuvat sen kovalevyn mukana.

Analoginen aineisto useimmiten on paremmin järjesteltyä. Ongelmaksi muodostuu tällöin se, että ei tiedetä, mitä tästä mahdollisesti vuosikymmenten aikajänteelle sijoittuvasta aineistosta pitäisi säilyttää. Arkistolaki velvoittaa ainoastaan julkishallintoa ja julkista tehtävää suorittavia organisaatioita, jolloin yksityisillä toimijoilla ei ole olemassa mitään yhtenäistä linjaa aineistojen säilyttämiseen. Samaan aikaan esimerkiksi yhdistysten ja seurojen aineistot muodostavat arvokkaan osan paikallista ja laajempaa kulttuurihistoriaa.

Ohjeistuksille on siis tarvetta: avoimesti ja helposti saatavat oppaat madaltavat oppimisen kynnyksiä ja voivat auttaa luomaan yhtenäisiä käytäntöjä. Joitakin ohjeistuksia toki jo löytyykään. Esimerkiksi Yksityiset keskusarkistot ry on omilla sivuillaan julkaissut kansantajuisen ohjeistuksen sähköiseen arkistointiin (Sähköisen tiedonhallinnan... 2018). Opas on kuitenkin perinteinen PDF-julkaisu ja myös hieman vaikeasti löydettävissä.

## **HELPOSTI YMMÄRRETTÄVISSÄ JA SAATAVILLA**

Hankkeessa kehitettävällä oppaalla oli lähtökohtaisesti kaksi tärkeää vaatimusta. Ensinnäkin sisällön pitäisi olla helposti luettavaa ja ymmärrettävää, koska opas on tarkoitettu kohde-ryhmille, joilla ei ole entuudestaan välttämättä mitään kosketusta digitointiin ja sähköiseen arkistointiin. Siksi digitoinnin salat oli avattava kansantajuisesti menemättä sisällön osalta liian syvälliseen tekniseen selostukseen.

Toiseksi ajatuksena oli, että siirrytään pois perinteisestä PDF-muotoisesta tekstistä ja laaditaan verkko-opas. Tällainen julkaisu mahdollistaa sellaisia tekniikoita, joita tavanomaisessa tekstijulkaisussa ei pystyisi toteuttamaan. Esimerkiksi pelillistämisen keinoja voidaan upottaa luontevasti osaksi kokonaisuutta. Digioppaassa lukijaa aktivoidaan leikkimielisillä "testaa tietosi" -tyyppisillä monivalintatesteillä. Sisältö on helposti hahmotettavissa etusivun näkymän perusteella. Sisältö on jaettu eri osioihin sen perusteella, mitä eri alueita digitointiin

ja sähköiseen arkistointiin liittyä. Kokonaisprosessin hahmottaminen on tärkeää, ja siksi yksi osio on omistettu digitointiprojektin toteuttamiselle alusta loppuun.

On aiheellista tarkentaa, että oppaan ei ole tarkoitus toimia Aku Ankka -tarinoista tutun “Sudenpentujen käsikirjan” tyyppisenä loputtomana tietolähteenä, mistä lukija löytää vastauksen jokaiseen kuviteltavissa olevaan digitoinnin ongelmaan. Esimerkiksi pelkästään tiedostomuodot ovat sellainen loputon suo, joista saisi kirjoitettua usean “Sinuhe egyptiläisen” verran tekstiä. Tämä ei olisi tarkoituksenmukaista, koska peruskäyttäjä ei todennäköisesti edes tarvitse tietoa siitä, millä tavalla jokin videokoodi eroaa jostakin toisesta, kunhan arkistoinnin kannalta parhaat tiedostomuodot esitellään ja kerrotaan niistä olennainen.

Tarkoitus on siis tarjota ruohonjuuritason digioppia koko kansalle. Oppaan avulla kuka tahansa ymmärtää, mitä on digitointi, mitä hyötyä siitä on, ja mikä on digiarkisto. Luonnollisesti myös arkistoinnin perusajatuksia ja käytäntöjä avataan, koska tavoitteiltaan digiarkisto ei eroa perinteisestä analogisesta arkistosta. Molemmat tähtäävät tiedon pitkäaikaisäilyttämisen turvaamiseen samoja periaatteita noudattaen.

## **VERKKO-OPPAAN TEKNINEN TOTEUTUS POHJAUTUU AVOIMEEN LÄHDEKODIIN**

Teknisesti sovellus on toteutettu uusimmilla tekniikoilla, ja se perustuu MVC-arkkitehtuuriin (Model-View-Controller). Back end eli palvelinpuoli on ohjelmoitu Node.js-tekniikoin käyttämällä Express.js -kehystä. Node.js mahdollistaa JavaScriptin suorittamisen palvelimella, ja se on helppo tapa käyttäjän syötteen käsittelemiseen. Esimerkiksi lomakkeella lähetettävät tiedot on helppo käsitellä Node.js-pohjaisella palvelimella. Avoimen lähdekoodin mahdollisimman tehokas hyödyntäminen ja jatkokehittäminen on ollut yksi Digitalian keskeisistä teemoista sovelluskehityksessä. Digioppaassa avoimen lähdekoodin tehokkuus on valjastettu käyttöön käyttöliittymän toteutuksessa, joka pohjautuu lähes kokonaan avoimeen lähdekoodiin. Kuva 1 esittelee oppaan päänäkymän.

## Digitointiopas

Digitaalinen Pimeä Aika lankeaa väistämättä yhteiskunnan ylle, jos emme nyt kääri kollektiivisia hihoja ja ala aktiivisesti säilöä kulttuuriperintöämme digiarkistoihin. Digitalia kantaa kortensa kekoon tarjoamalla Kansalaisyhteiskunnan pitkäaikaissäilytysoppaan, joka toimii digiarkistoinnin tietopakettina.



*Kuva 1. Digioppaan etusivu työpöytäkoneella*

Käyttöliittymä on toteutettu HTML5-tekniikoilla käyttämällä Bootstrap-kehystä ja erilaisia vapaan lähdekoodin JavaScript-kirjastoja. Käyttöliittymä on suunniteltu responsiiviseksi, eli se mukautuu luontevasti sekä mobiililaitteiden että työpöytäkoneiden eri kokoisille näytöille. Mobiilinäkymässä etusivun ”kortit” asettuvat yhteen sarakkeeseen päällekkäin, mikä tekee selaamisesta helppoa.

Oppaan kohderyhmiä on useita, eli sitä ei ole kohdennettu millekään tietylle toimialalle tai käyttäjäryhmälle. Näissä puitteissa pystyimme itse toteuttamaan sisällön sen perusteella, minkä katsoimme vastaavan taustatutkimuksessa esille nousseita tarpeita. Tasaisin väliajoin pidettiin katselmuksia ja sisältöä kehitettiin eri tahoilta saadun palautteen perusteella. Ensimmäisten palautteiden perusteella huomiota kiinnitti oppaan liiallinen luettelomaisuus. Kunnianhimoisista tavoitteista huolimatta sisältö siis muistutti alussa tyypillistä oppikirjatekstia, jossa sisältö ei niinkään rakennu prosesseiksi, vaan siinä luetellaan väitteitä. Tämä ei ole oppimisen kannalta tehokasta, koska tekstiä täytyy ”päntätä” ja opetella ulkoa (Aalto 2007).

Palautteen perusteella opasta muutettiin helpommin luettavan verkkotekstin muotoon esimerkiksi lyhyemmillä kappaleilla ja nostoilla. Yhteistyökumppaneilla oli paljon hyviä ehdotuksia, joiden perusteella sisältöä on kehitetty. Esimerkiksi digitointiprojektin ymmärtäminen ja osaaminen kokonaisuutena oli yksi tärkeä seikka, joka palautteen perusteella haluttiin tuoda

paremmin esille. Jokaisen iterointikierroksen jälkeen opas muuttuikin hieman - ja tulee jatkossakin muuttamaan muotoaan. Toisin kuin arkistoitu tieto, opas elää ja kehittyy käyttäjäkunnan tarpeiden, toiveiden ja palautteen mukaan. Tämä on samalla yksi verkkojulkaisun eduista, koska sisältöä ja tarvittaessa myös ulkoasua pystytään muuttamaan nopeastikin.

## **ARKISTOINNIN VASTUU SIIRTYY SUKUPOLVELTA TOISELLE**

Kriittisimmät ovat puhuneet digitaalisesta pimeästä ajasta tai digitaalisesta keskiajasta (Digital Dark Age, Cerf 2015). Termillä viitataan digitaalisessa muodossa tallennetun kulttuurihistoriallisesti merkittävän tiedon katoamiseen. Yhä enemmän ja enemmän viestinnässä nojaamme syntysähköiseen aineistoon, jota ei kuitenkaan säilytetä. Sähköpostilaatikoita tyhjenetään tasaisin väliajoin, puhelimia vaihdetaan tiuhaan ja viestihistoriat katoavat, ja otamme valokuvia älypuhelimilla, vaikka kuvien säilymistä ei kuitenkaan varmisteta. Lisäksi tiedostomuodot ja ohjelmistot vanhenevat, eikä digitaalisia tiedostoja välttämättä pystytä enää tulevaisuudessa avaamaan. Jos viittaamme kintaalla pitkäaikaissäilyttämisen tärkeydelle, niin pian saatammekin huomata, että digitaalinen jalanjälkemme katoaa eikä digiajan jälkipolville jää meistä kollektiivista identiteettiämme määrittelevää tietoa – ainoastaan tomua ja tuhkaa.

Tämä kaikki voidaan välttää, kun tartumme tuumasta toimeen ja alamme omalla toiminnallamme turvata tiedon pitkäaikaista säilymistä. Tiedon tallentaminen arkistointiin soveltuvilla tiedostomuodoilla ja säilöminen oikeille tallennusmedioille ei yksin riitä, vaan lisäksi täytyy aktiivisesti huolehtia tiedon säilymisestä. Olennaista on jatkuva tiedon käytettävyyden, saatavuuden ja ymmärrettävyyden turvaaminen. Myös digiaikana arkistoinnin vastuu siirtyy sukupolvelta toiselle.

## LÄHTEET

Aalto, E. 2007. Hyvä oppimateriaali – autenttista toimintaa. WWW-dokumentti. Päivitetty 23.11.2007. Saatavissa <http://www.edu.fi> [viitattu 8.5.2019].

Cerf, V. 2015. Google's Vint Cerf warns of 'digital Dark Age'. WWW-dokumentti. Päivitetty 13.2.2015. Saatavissa: <https://www.bbc.com/news/science-environment-31450389> [viitattu 10.4.2019].

Sähköisen tiedonhallinnan ja arkistoinnin ohjeet yksityisaineistoille. 2018. Yksityiset keskusarkistot ry. WWW-dokumentti. Päivitetty 25.4.2018. Saatavissa: <http://www.yksityisetkeskusarkistot.fi/> [viitattu 9.4.2019].



# DIGITAALISTEN AINEISTOJEN SISÄLTÖKÄYTTÖTOIVEITA TUTKIJOILTA

Tuula Pääkkönen, FM, tietojärjestelmäasiantuntija, Kansalliskirjasto  
& Johanna Lilja, FT, palvelujohtaja, Kansalliskirjasto

Digitaaliset aineistot käyttöön -hankkeessa yhtenä taustateemana on alkuperäisen suunnitelman mukaisesti ollut Kansalliskirjaston digitaalisten aineistojen käytön edistäminen ja innovaatioaihioiden kehittäminen niistä sellaisissa yrityksissä ja yhteisöissä, joiden historiaan aineistot liittyvät (Lilja & Hakkarainen 2018). Käyttäjien toiveet kilpistyvät kuitenkin usein tekijänoikeuslain sekä Kansalliskirjaston ja tekijänoikeusjärjestöjen välisten, aineiston käyttöä säätelevien sopimusten asettamiin rajoitteisiin. Tässä artikkelissa käymme läpi projektin aikana ja osin aiemminkin saatuja aineistojen käyttöoikeuteen liittyviä palautteita ja toiveita eri käyttäjäryhmiltä.

## KANSALLISKIRJASTON AINEISTOT JA NIIDEN KÄYTTÖOIKEUDET

Kansalliskirjastolla on tekijänoikeuslain suoma mahdollisuus digitoida vapaakappaleina saatuja, kansalliskokoelmaan kuuluvia julkaisuja säilytystarkoituksiin (TekL 16§). Suurimmat digitoidut aineistokokonaisuudet digi.kansalliskirjasto.fi -palvelussa koostuvat sanoma- ja aikakauslehdistä, joita on kattavasti digitoitu ensimmäisestä sanomalehdestä 1771 vuoteen 1935 (Arpiainen 2019). Lisäksi muutamia lehtinimekkeitä on digitoitu pitemmältäkin ajalta, osa niiden alusta nykypäivään. Lähtökohtaisesti tekijänoikeuden alaista vapaakappaleina luovutettua aineistoa saa käyttää vain Kansalliskirjastossa ja muissa vapaakappalekirjastoissa laitteilla, joissa digitaalisten jäljenteiden ottaminen on estetty (Kansalliskirjasto 2016).

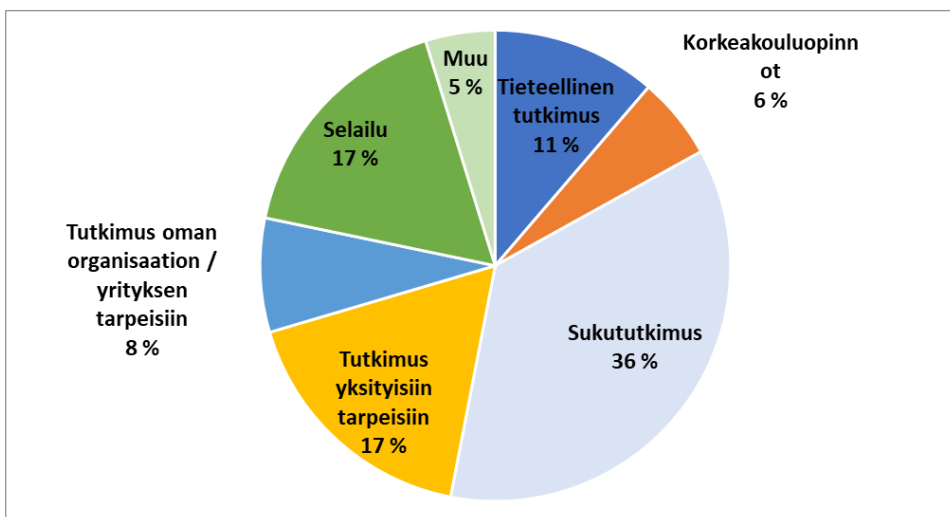
Digi.kansalliskirjasto.fi -palvelussa tarjottavien sanoma- ja aikakauslehtien käyttöoikeudet vaihtelevat riippuen niiden julkaisuajasta ja niihin liittyvistä sopimuksista. Tekijänoikeus raukeaa, kun 70 vuotta tekijän kuolemasta on kulunut (TekL 43§). Käytännössä sataa vuotta vanhemmat lehdet eivät välttämättä ole kokonaan tekijänoikeuksista vapaita, mutta riski väärinkäytöksille on hyvin pieni. Erityisesti sanoma- ja aikakauslehdissä monimutkaisuutta lisää se, että jokaisen niteen, jokaisen sivun, jokaisella jutulla ja kuvituksilla on omat tekijänsä (Kansalliskirjasto 2019a).

Tekijänoikeuslakiin perustuva sopimuslisenssikäytäntö mahdollistaa Kansalliskirjastolle ratkaisun, jossa tekijänoikeusjärjestön kanssa sovitaan tekijänoikeudenalaisen aineiston avaami-

sesta (TekL 16 d§, OKM 2019). Ensimmäinen sopimus solmittiin Kopioston kanssa vuonna 2017 ja sillä voitiin avata kaikille käyttäjille vuosien 1917-1929 sanoma- ja aikakauslehdet. Tätä sopimusta jatkettiin syksyllä 2018, ja se on voimassa vuoden 2020 loppuun. Sopimus sallii aineiston haun, selailun ja omien leikkeiden tekemisen palveluun, mutta ei aineiston julkaisemista muualla. Toinen, ehdoiltaan jossain määrin erilainen sopimus, solmittiin Kopioston kanssa vuosien 1930-2010 digitoitujen lehtiaineistojen avaamiseksi tutkimuskäyttöön (Mäkelä 2018). Tämä sopimus on luonteeltaan pilottisopimus. Erillistä kirjautumista ja tutkimuskäytön vahvistavaa kyselylomaketta vaativa avaamisratkaisu on käytössä kuudessa yliopistossa vuoden 2019 loppuun. Tässä sopimuksessa määritelty tutkimuskäyttö kattaa monipuolisemman käytön kuin edellä mainittu vuosien 1917-1929 lehtiaineiston yleisökäyttö. Sopimus mahdollistaa mm. tekstintalouhinnan lehtien sisällöstä ja sen kautta syntyvien datapakettien tallentamisen Tietoarkistoon (Mäkelä 2018). Kansalliskirjasto on sopinut Kopioston kanssa myös joitakin suppeampia aineistokohtaisia sopimuksia, joista on listaus [digi.kansalliskirjasto.fi](http://digi.kansalliskirjasto.fi)-palvelun etusivulla.

## KETKÄ SISÄLTÖJÄ KÄYTTÄVÄT?

Digi.kansalliskirjasto.fi -palvelun käyttäjäkunta on laaja ja kasvaa jatkuvasti. Vuonna 2018 palvelusta tehtiin 12,7 miljoonaa latausta (Kansalliskirjasto 2019b). Käyttäjryhmiä on hyvin moninaisia. Vuonna 2018 tehdyssä käyttäjäkyselyssä erilaiset tutkijat eri rooleissaan olivat tärkeimmät käyttäjäryhmät, mutta myös selailijat olivat oma käyttäjäpopulaationsa (Pääkkönen & Kettunen 2018). Kuva 1 näyttää tietoja käyttäjien ilmoittamasta käyttötarkoituksesta. Käyttäjän oli mahdollista valita yksi tai useita tarkoituksia.



Kuva 1. Digi.kansalliskirjasto.fi – palvelun käyttäjien käyttötarkoituksia (2018)

## KÄYTTÖESIMERKKEJÄ

Hankkeen aikana Kansalliskirjastolle on tullut useita aineiston käyttöön liittyviä kysymyksiä. Tekijänoikeuskysymykset ovat yleensä monimutkaisia ja vastaukset vaihtelevat riippuen mm. käyttötarkoituksesta, aineistosta, sen iästä ja aineiston käyttöä rajaavista sopimuksista. Tarkastelemme seuraavassa neljää aineiston käyttöön liittyvää kysymystä.

### Kysymystyyppi 1 – Miksi aineistoa ei avata?

Yleisin Kansalliskirjaston kyselyissä, verkkopalautteena ja sosiaalisessa mediassa saama palaute liittyy siihen, miksi vain rajattua osaa aineistosta pääsee lukemaan. Miksei avata kaikkea digitoitua aineistoa, kun sille olisi niin paljon hyödyllistä käyttöä?

Eräs käyttäjä lähetti pitkän kysymyksen jo projektin alkuaikoina, jolloin aineistoa ei vielä ollut sopimuslisenssillä avattu. Palautteen antajan luvalla osa palautteesta näkyy kuvassa 2.

En ymmärrä, miksi säädökset edelleen estävät vuodet 1911-1916 (ja siitä eteenpäin). 100 vuotta lienee jonkinlainen henkilösuojaraja, mutta miksi se on jäädytetty vuoteen 1910? Toisaalta juuri nämä vuodet ovat nyt mielenkiintoisia, koska Suomen itsenäisyys täyttää 100 vuotta ensi vuonna. Juuri 1910-luvulla eri puolilla maata tapahtui huiman paljon konkreettisia asioita, joista juuri sanomalehdet kertoivat. Paikallishistoriasta kirjoittaminen ja esitelmöiminen tältä kaudelta on mahdotonta tai ainakin hyvin hankalaa. Tulee esitettyä näennäistietoja. Tai tulee esitettyä professoreiden ja tutkijoiden opettelevien huolimattomasti hutaistuja yleistyksiä aikakaudelta. (Luulevat, että historia tapahtuu yleisesti.)

Toki lehtiä voi mennä penkomaan arkistoihin ja joihinkin kirjastoihin, jos sattuu pääsemään ja ehtimään odottelemaan aineiston saapumista. Maseudulta käsin tilanne on toivoton; on matkustettava kaupunkiin, josta vapaakappaletoimisto löytyy. Pitää heti pienellekin asialle varata viikko. ("Vapaakappaleasemien aineistoja voi lukea, kuunnella ja katsella. Sisällöstä ei voi valmistaa digitaalista kopiota.") Tiedonsaannissa on nyt selkeä pullonkaula!

*Kuva 2. Käyttäjäpalautetta tekijänoikeusasioista*

Sama palaute toistuu tuoreimmassa asiakastytyväisyyskyselyssä useaan otteeseen, milloin enemmän, milloin vähemmän kärjekkäästi: ”Palvelu toimii oikein hienosti, mutta sen tarjonnassa sisällössä hämmästyttää suuresti nämä aineistojen rajoitukset, 1918-1929 vain tämän vuoden ajan, vuodesta 1930 eteenpäin vain tietyillä työpäätteillä, jne. Suunnilleen yhtä suuresti kiinnostaisi tietää näiden rajoitusten perusteet, niitä ei missään mainita.” (Pääkkönen & Lilja 2018)

Käytön laajentamiseen liittyvä palaute viittaa siihen, että verkkokäyttäjät ovat tottuneet siihen, että kaikki on saatavilla, eivätkä tunne tekijänoikeuden asettamia rajoitteita, vaikka [digi.kansalliskirjasto.fi](http://digi.kansalliskirjasto.fi) -palvelussa käyttöehdot on selkeästi linkitetty kaikkiin aineistoihin. Harva hyväksyy vastausta, että kaikki digitaalinen aineisto on käytettävissä vapaakappalekirjastoissa niiden aukioloaikoina.

## Kysymystyyppi 2 - Missä voin esittää aineistoa toisille?

Lehtiaineisto on värikästä, usein hauskaa, ja siksi monet haluaisivat värittää sillä esityksiään. Yksi kysymystyyppi liittyykin aineistojen esittämiseen muille.

Tekijänoikeuslaki rajoittaa tekijänoikeuden alaisen aineiston välittämisen digitaalisena vain vapaakappalekirjastoihin. On hyvä huomata, että näyttäminen, esittäminen ja välittäminen ovat erillisiä käytöntapoja ja vaativat erilliset käyttöoikeudet (TekL 2§). Esimerkiksi vapaakappalekirjaston tiloissa on mahdollista välittää aineistoa käyttäjille, mutta muualla – vapaakappalekirjaston seinien ulkopuolella – tämä ei kävisi, ellei oikeudenhaltijoiden kanssa ole tästä erikseen sovittu. Aineistoa voidaan myös siteerata, jos aineistoa siteerataan uudessa (ilmaisullisessa) teoksessa.

## Kysymystyyppi 3 - Saako aineistoa jäljentää? - "Saisinko tämän artikkelin digitaalisena"

Tekijänoikeuksista vapaata aineistoa käyttäjät voivat kopioida ja jakaa mielin määrin. Sopi-  
muksin avattua aineistoa, eli yleisökäytössä olevia lehtiä 1917-1929 ja tutkijakäytössä olevia lehtiä 1930-2010, saa myös kopioida yksityiseen käyttöön (Kansalliskirjasto 2019a). Kopioita ei kuitenkaan saa edelleenjulkaista eikä jakaa. Tämä on osoittautunut oppimismahdollisuudeksi hankkeessa, jossa lehtien sisällölle on keksitty uusia käyttötarkoituksia, mm. yritysten mainonnassa tai mainostuotteissa. On ymmärrettävää, että käyttäjästä tuntuu harmilliselta, että hyvä kuva-aiho jää käyttämättä. Ongelmalliseksi tilanteen tekee se, että kuvaajat, kuten myös uutisten ja ilmoitusten tekijät ovat usein anonyymeja, eikä oikeudenhaltijan kanssa sopimiseen siksi ole välineitä.

Vapaakappalekirjastojen asiakkaiden joukossa hämmästyttäviä aiheuttaa se, että vapaakappaletyöasemilla tarjotusta digitaalisesta aineistosta ei saa digitaalista kopiota. Usein käyttäjät tahtoisivat tallettaa sähköisenä löytämänsä sisältönsivun. Kuitenkaan tämä ei ole mahdollista, sillä laitteistoihin on tarkoituksellisesti rakennettu tekniset estot, joista työasemia käyttävät eivät yleensä ole ensimmäisillä käyttökertoillaan tietoisia. Vapaakappalekirjastoja on Suomessa 6 (Pitkälä 2019), joissa on yhteensä käytettävissä noin 20 vapaakappaletyöasemaa. Niiden käyttäminen on mahdollista kirjaston aukioloaikoina. Teknisten rajoitusten toteuttaminen vaatii kirjastoilta voimavaroja.

Useimmat maat painiskelevat saman ongelman kanssa - kysyntää digitoiduille lehdille olisi enemmän kuin on mahdollista tarjota. Erilaisin sopimuksin on kuitenkin onnistuttu avaamaan lehtiä. Esimerkiksi Itävallassa sanomalehdet ovat kaikille kansalaisille saatavilla vuoden 1948 loppuun asti noudattaen sitten liikkuvaa 70 vuoden avausperiaatetta (Austrian National Library 2019). Tämän mahdollistavat sekä sopimukset, tekijöiden selvittäminen digitoinnin jälkeen ja se, että taktiikkana on toimintansa lopettaneiden kustantajien aineistojen avaaminen. Näillä keinoin hallitaan erilaisia riskejä.

#### **Kysymystyyppi 4. Voinko hyödyntää aineistoa datana? - Voinko saada datapaketin?**

Kuten usein todetaan, data on uusi öljy. Lehtiaineiston käyttökään ei rajoitu painettujen lehtien selailuun ja lähilukuun, vaan moni toivoo saavansa aineiston käyttöön datana, eli kuvasta OCR-prosessin kautta tuotettuna tekstimassana, jota tutkija voi käsitellä koneellisesti. Avoimen tieteen periaatteiden mukaisesti tutkijan pitäisi tallettaa ja avata muille tutkijoille raakadatasta työstämänsä datapaketti, jotta sitä voidaan tarvittaessa käyttää uudelleen. (Fält 2016, 39-48)

Lehtiaineiston tutkimuskäyttöä pilotoivaan sopimukseen sisältyy käyttäjän oikeus tehdä datapaketteja ja tallettaa ne Tietoarkistoon. Datan tutkimuskäyttö ei edellytä sopimista, koska tekijänoikeus ei ulotu dataan tai faktoihin, vain muotoon. Jos louhinta onnistu ilman kappaleiden valmistamista, lupaa ei tällöin tarvita. Oikeudellisten kysymysten lisäksi datan louhimiseen liittyy myös teknisiä haasteita.

#### **Mitä työkaluja voisin aineistojen kanssa käyttää?**

Liittyen aineistojen saatavuuteen seuraava kysymys on, miten ja millaisin työkaluin laajoja aineistomassoja voi hyödyntää. Osa tutkijoista käyttää vieläkin paperia, osa ohjelmoi ja osa hyödyntää hiljalleen tutkijaekosysteemiin tulleita palveluita, jotka helpottavat aineistojen perusanalyysiä ja jättävät tutkijalle tutkimuskysymyksen ja varsinaisen tutkimustyön aspektien pohdinnan.

Kansalliskirjasto on tehnyt uudistettua tutkijaohjeistusta ja hankkeen puitteissa olemme tehneet myös ohjekirjan <https://github.com/natlibfi/digitalia-notebook>, jossa kädestä pitäen käydään python-ohjelmointikielen esimerkkien avulla läpi, miten kukin itse voisi digitaalisia aineistoja hyödyntää. Näiden esimerkkien myötä on mahdollista tutustua käyttötapoihin ja laajentaa ohjelmallista hyötykäyttöä omiin tarkoituksiin.

## **YHTEENVETO**

Digi.kansalliskirjasto.fi -palvelussa oleva aineisto syntyy digitointiprosessissa, joka tuottaa päivässä tuhansia sivuja tekstiä - osa vapaata, osa tekijänoikeuden alaista. Aineisto kuvaillaan lehtinimekkeiden, volyymien ja numeroiden tasolla, mutta artikkeli- tai kuvakohtaisten tekijänoikeustietojen antaminen käyttäjälle on mahdotonta. Käyttöehdot voi tarjota vain yleisellä tasolla nimeke- ja ajanjaksokohtaisesti.

Lehtiaineistojen käyttöoikeuksiin liittyvät palautteet ja kysymykset osoittavat sen, että tekijänoikeudellisten asioiden ymmärrettäväksi tekeminen käyttäjille on vaikeaa. Toisaalta palautteet osoittavat myös sen, että aineistolla olisi valtava käyttöpotentiaali. Käytön mahdollistaminen lisäisi kulttuuriperintöaineistojen monipuolista hyödyntämistä niin tutkimuksessa, yksityisessä harrastekäytössä kuin yritystoiminnassakin.

Tekijänoikeuskysymykset ovat asia, joka kaikkien aineistoja hyödyntävien olisi tiedostettava. Yhdessä kulttuuriperintöorganisaatiot, tekijöiden oikeuksia ja etuja ajavat tekijänoikeusjärjestöt ja tutkijat voisivat luoda malleja, joilla aineiston tutkimuskäyttö onnistuisi. Tutkimuksen vaikuttavuuden kasvattamisen takia tutkijoiden tulisi pystyä kommunikoimaan tutkimuksestaan, koska siitä hyötyy koko yhteiskunta, kun päästään asioiden juurisyihin.

Toivokaamme siis, että tuoreeseen tekijänoikeusdirektiiviin perustuva kansallinen lainsäädäntö tukee mahdollisimman hyvin tutkijoiden käytötapauksia. Kannattaa seurata tilanteen kehitystä ja antaa lausuntoja, kunhan työ käynnistyy. Kommentteja eri sidosryhmiltä tarvitaan pikapuoliin.

## **KIITOKSET**

Kansalliskirjaston Pekka Heikkiselle arvokkaista kommentteista tekstin sisältöön.

## LÄHTEET

Arpiainen, H. 2019. Digitoimme nyt (2018). WWW-dokumentti. Saatavissa: <https://www.kansalliskirjasto.fi/fi/digitoimme-nyt> [viitattu 20.8.2018].

Austrian National Library. 2019. ANNO - AustriaN Newspapers Online. WWW-dokumentti. Saatavissa: <http://anno.onb.ac.at/> [viitattu 29.4.2019].

Finlex 2017. Tekijänoikeuslaki 404/1961. Noudettu osoitteesta <http://www.finlex.fi/fi/laki/ajantasa/1961/19610404#L3P40b>

Fält, K. 2016. Tutkijoiden digitaaliset tutkimusaineistot ja datan avoin jakaminen digitaalisen humanismin kontekstissa. Teoksessa Elo, K. (toim.), Digitaalinen humanismi ja historiatieteet. Turku: Turun historiallinen yhdistys 2016, 39-75.

Kansalliskirjasto 2016. Elektronisten aineistojen käyttöoikeudet ja käyttöehdot. WWW-dokumentti. Saatavissa: <https://www.kansalliskirjasto.fi/fi/aineistot/aineiston-saatavuus-ja-kaytto/elektronisten-aineistojen-kayttooikeudet-ja-kayttoehdot> [viitattu 30.4.2019].

Kansalliskirjasto 2019a. Kansalliskirjaston Digi - FAQ. WWW-dokumentti. Saatavissa: <https://digi.kansalliskirjasto.fi/faq> [viitattu 30.4.2019].

Kansalliskirjasto 2019b. Kansalliskirjaston vuosi 2019. Noudettu osoitteesta <http://www.doria.fi/handle/10024/168085>.

Lilja, J. & Hakkarainen, J.-P. 2018. Re-defining our services - National Library's new initiatives to support open science. Helsingin Summit 2018, Helsinki.

Mäkelä, E. 2018. Sanoma- ja aikakauslehtiaineiston tutkimus- ja opetuskäytön edistäminen. WWW-dokumentti. Saatavissa: <https://www.kansalliskirjasto.fi/fi/projektit/sanoma-ja-aikakauslehtiaineiston-tutkimus-ja-opetuskaayton-edistaminen> [viitattu 31.10.2018].

OKM 2019. Sopimuslisenssit. WWW-dokumentti. Saatavissa: <https://minedu.fi/sopimuslisenssit> [viitattu 30.4.2019].

Pitkälä, M. 2019. Vapaakappaletoimisto. WWW-dokumentti. Saatavissa: <https://www.kansalliskirjasto.fi/fi/vapaakappaletoimisto> [viitattu 14.4.2019].

Pääkkönen, T. & Kettunen, K. 2018. Kansalliskirjaston sanomalehtiaineistot: käyttäjät ja tutkijat kesällä 2018. Informaatiotutkimuksen päivät 2018. Noudettu osoitteesta <http://urn.fi/URN:NBN:fi-fe2018110247067>.

Pääkkönen, T. & Lilja, J. 2018. Hieno palvelu, mutta sisältöä lisää – Kansalliskirjasto kyseli Digi-palvelun käyttökokemuksia. Tietolinja 2/2018. Noudettu osoitteesta <http://urn.fi/URN:NBN:fi-fe2018092336401>.



# HISTORIALLISET LEHDET YRITYKSIÄ INSPIROIVANA TARINA-AINEISTONA

Torsti Hyryläinen, HT, tutkimusjohtaja & Manu Rantanen, FM, projektipäällikkö & Juuso Muttilainen, LuK, projektityöntekijä, Helsingin yliopisto, Ruralia-instituutti

Pohdimme artikkelissa, miten Kansalliskirjaston digitaalisen kirjaston verkossa julkaisemaa digitaalista aineistoa (<https://digi.kansalliskirjasto.fi>) voidaan hyödyntää yritysten tuotteiden ja palvelujen kehittämisessä, markkinoinnissa ja mainekuvan rakentamisessa. Havainnot ja tulokset perustuvat Helsingin yliopiston Ruralia-instituutissa 2017–2019 välisenä aikana koottuun 77 yritystä ja yhteisöä koskevaan digitaaliseen artikkeliaineistoon. Analysoimme siitä tarkemmin 25 yksittäistä yritystapausta ja lopulta kaikkiaan 13 yrityksen kanssa toteutettiin tapaaminen, jossa kehitelimme ja ideoimme yhdessä yrittäjien kanssa aineiston luovaa käyttöä.

Kaikki lähestymämme yritykset pitivät aineistoa kiinnostavana. Aineisto nähtiin hyödyllisenä yritysten tarinankerronnan kannalta ja siitä löytyi myös runsaasti aineksia tuote- ja palveluinnovaatioiksi monien eri toimialojen yrityksille. Merkittävänä rajoitteena yritykset kohtasivat tekijänoikeudelliset kysymykset, jotka vaikeuttavat erityisesti kuvien käyttöä. Erityisesti pienillä yrityksillä on myös rajalliset resurssit perehtyä syvällisesti historiallisiin lehtiaineistoihin. Digitaalinen aineisto osoittautui todelliseksi aarreaksiksi ja inspiraation lähteeksi. Sanoma- ja aikakauslehdissä esitetyt tarinat, kuvitus ja näkökulmat ovat kiistatta sekä historiallisen ymmärryksen että uusien innovaatioiden lähde.

Avoimen digitaalisten aineistojen käyttöä on aiemmin Etelä-Savon kontekstissa tarkasteltu Rajahongan ja Villmanin (2018) raportissa. Siinä painotus oli avoimen datan käyttömahdollisuuksien esittelyssä ja hyödyntämisen potentiaalın tarkastelussa. Avoimen datan liiketoiminnallisessa hyödyntämisessä eteläsavolaiset yritykset ovat Kaija Villmanin arvion mukaan (emt., 8) vasta hyvin alussa. Tässä artikkelissa keskitymme puolestaan yhteen avoimen datan aineistolähteeseen ja tarkastelemme lähemmin sen erityispiirteitä ja käyttömahdollisuuksia.

## JOURNALISTISET TEKSTIT UUSIEN IDEOIDEN LÄHTEENÄ

Vanhat, 1770-luvulta lähtien ilmestyneet suomalaiset sanoma- ja aikakauslehdet ovat erityistä historiallista aineistoa. Niiden luomaa kerrontaa on luonnehdittu historian ensimmäiseksi vedokseksi. Journalismin on sanottu esittävän sitä, millä on merkitystä nykyhetkessä. Sanoma- ja aikakauslehtien kuvauksia voi luonnehtia myös tapahtumien päiväkirjamaiseksi dokumentoinniksi. Lehtijournalismi voi sisältää virheitä ja epätarkkuuksia, mutta samalla sanoma- ja aikakauslehdet hahmottavat laajempaa ajankuvaa, millaista elämä on ollut ja

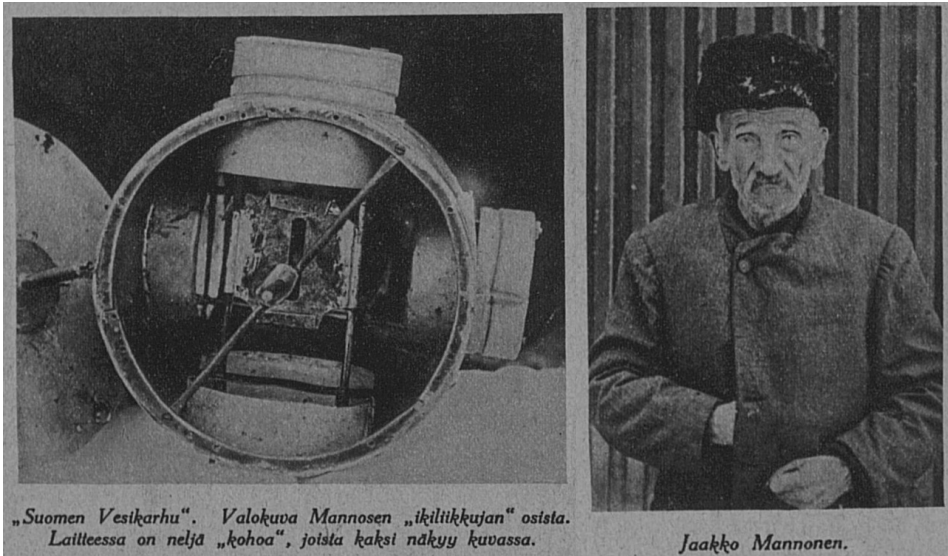
millaisena se on kulloinkin nähty. Kiinnostavaa ei kuitenkaan ole vain journalistisen esityksen historiallinen totuudellisuus tai sisältö, vaan erityisesti se, miten journalismi luo merkityksiä historian virrassa. On myös hyvä muistaa, että lehdillä on jokin toimituksellinen tai journalistinen linja, ja näkökulma asioihin on sen mukainen. Samasta tapahtumasta eri lehdet voivat kertoa erilaisen tarinan. (ks. Seuri 2014; Kitch 2008)

Lähestymistapamme kannalta erityisen kiinnostava on lehtijournalismin rooli yhteiskunnallisten tapahtumien valikoivana raportoijana. Tällä tarkoitamme ensinnäkin sitä, että juuri sanoma- ja aikakauslehtiartikkelien kautta tulee esiin se, mikä asia kulloinkin nousee ”uutiseksi”, eli mikä tekee asiasta tai tapahtumasta niin huomattavan, että se nostetaan otsikoihin ja halutaan jakaa lukijakunnalle merkittävänä tietona. On myös paikallaan todeta, että journalistisessa esitystavassa on vuosisatojen ja -kymmenten saatossa tapahtunut erittäin suuri muutos. Ennen asioista uutisoitiin usein tavalla, joka nykypäivän näkökulmasta tuntuu oudolta, huvittavalta tai jopa mahdottomalta.

Journalistien ikiaikainen tehtävä ”uutisoida” sitä, mikä voisi muuten jäädä pimentoon, on erityisen kiinnostava piirre sanoma- ja aikakauslehtikirjoittelussa. Lehdet myös kokoavat yhteen ja poimivat ajankohdalle tyypillisiä näkemyksiä ja ajatuksia (ks. Alasuutari 2017). Journalistilla voi olla erilaisia sidoksia, mutta toisaalta hän on myös vapaa valitsemaan kohteensa ja näkökulmansa. Nämä historialliset kuvaukset maailmasta voivat avata nykypäivän ihmisille uuden tavan katsoa sekä mennyttä että nykyisyyttä.

Historiallisten sanoma- ja aikakauslehtien yrityksiä koskeva uutisointi voi olla monella tavalla inspiroivaa. Kirjoitukset ja kuvat voivat ”etäisyyden päästä” katsoen tuoda yrityksen nykyisyyteen aineksia, jotka tässä ajassa tulkittuina luovat uusia merkityksiä, tunnetiloja tai ideoita, joita ei muuten synny. Näin ajateltuna historiallinen lehtiaineisto tarjoaa aiemmin hyvin vähän hyödynnetyn mahdollisuuden yritysten tuotteiden ja palvelujen kehittämiseen ja erityisesti niiden tarinallistamiseen.

Historiallisten lehtien tyyli poikkeaa usein paljonkin nykyisestä tavasta. Aiheet ja kerronta vaikuttavat nykyhetkessä tarkasteltuina koomisiltakin, vaikka ne omana aikanaan ovat olleet vakavasti otettavia uutisia (kuva 1).



*Kuva 1. Ikiliikkujan keksimisestä kertovan uutisen kuvitusta vuodelta 1925 (Suomen Kuvalehti, 21.03.1925, nro 12, s. 13)*

Kuvaamme tässä artikkelissa sitä, missä määrin edellä kuvattu ajatus historiallisten sanoma- ja aikakauslehtien käytöstä yritysten inspiraation lähteenä toteutui, kun kävimme Kansalliskirjaston historiallista sanoma- ja aikakauslehtiaineistoa läpi eteläsavolaisten yritysten kanssa.

## **DIGITAALISTEN AINEISTOJEN LOUHINTA**

Kuvaamme seuraavassa aineiston kokoamisen ja analysoimisen prosessin. Siihen on hankkeen aikana osallistunut lisäksi kolme projektityöntekijää (Ville Malila, Joanna Asumus ja Annika Manner). Kokonaisuutena prosessi on esitetty kuviossa 1.

	VAIHE 1 KARTOITUS	VAIHE 2 LOUHINTA	VAIHE 3 KONTAKTOINTI	VAIHE 4 YHTEISKEHITTÄMINEN	VAIHE 5 HAVAINNOT/ TULOKSET
Tarkasteltavien yritysten ja yhteisöjen määrä	Etsitään potentiaaliset yritys- ja yhteisökohteet sekä paikat Etelä-Savosta.	Louhitaan kohdekohtaista tietoa keskittyen aineistoista kumpuaviin tarinoihin.	Valitaan löydettyjen materiaalien kannalta kiinnostavia yrityksiä ja yhteisöjä.	Esitellään yrityksille digitaalisten aineistojen käyttöä.	Raportoidaan tulokset ja kuvataan prosessi.
	Etsitään erilaisia kohteisiin liittyviä hakusanoja.	Kohteista etsitään kiinnostavaa ja mahdollisesti toimijoita inspiroivaa materiaalia ja aineisto-esimerkkejä.	Otetaan yhteyttä kohteisiin, ja tarjotaan aineiston esittelyä.	Käydään keskustelua löydetystä materiaalista sekä kehitetään yhdessä yritysten kanssa aineiston hyödyntämisen tapoja.	
	Eteläsavolaiset yritykset ja yhteisöt	77 yritystä ja yhteisöä	25 yritystä ja yhteisöä	13 yritystä	
	2017 ->	Prosessin eteneminen	2018	2019 ->	

*Kuvio 1. Prosessin vaiheittainen eteneminen aineiston kokoamisesta yhteiskehittelyyn*

Aineiston louhinta aloitettiin etsimällä hakusanoiksi potentiaalisten yritys- ja yhteisökohteiden nimiä sekä niihin liittyviä henkilöitä ja paikkoja Etelä-Savosta. Tietoa haettiin yritysluette-loista ja historiikeista pitkään toimineista yrityksistä, jotka olisivat edelleen toiminnassa ja joista näin voisi löytyä osumia historiallisista aineistoista. Lisäksi kartoista etsittiin paikkojen ja järvien nimiä, jotka liittyvät tunnettuihin yrityksiin.

Tämän lisäksi etsittiin eri aloihin liittyviä hakusanoja. Tarkoituksena oli löytää mm. perin-neruokiin ja matkailuun liittyvillä aloilla toimiville yrityksille tai yhteisöille aineistoa edelleen hyödynnettäväksi. Hakusanojen haarukoinnissa hyödynnettiin Ruralia-instituutin ruoka-alan sekä Mikkelin kehitysyritys Miksei Oy:n asiantuntijoita, joita haastateltiin aiheeseen liittyen. Louhinnassa hyödynnettiin digi.kansalliskirjasto.fi -sivuston hakutyökaluja sekä Boolean operaattoreita. Lisäksi käytettiin niin sanottua sumeaa hakua, joka toi esiin myös hakusanaa muistuttavia sanoja. Tämä oli tarpeen erityisesti vanhemman, fraktuuralla kirjoitetun aineis-ton haarukoinnissa. Vaikka monet haut johtivat harhaan, kuten samannimisiin henkilöihin muualla maassa, löydettiin runsaasti erilaisia tarinanaihioita ja yritysmateriaalia. Uutisten ja kuvausten lisäksi ajankuvaa avasivat muun muassa varhaiset mainokset (kuva 2).

**Parhaiden lajiteltu**  
**warasto**  
**Seinä-, tasku-, pöytä-**  
**Kelloja**  
 kellonperä, formuusia, gramofoneja,  
 ilmapuntareita, lämpömittareja, filmä-  
 laseja, sähkötaskulamppuja y. m. y. m.  
 halvalla  
**Oskar Laitisella,**  
**Mitteli. Kirkkokatu 54.**

*Kuva 2. Kelloliikkeen mainos vuodelta 1908 (Vapaus, 07.11.1908, nro 125, s. 4)*

Alkuvaiheessa hakuja toteutettiin useamman projektityöntekijän voimin. Selvyiden vuoksi jokainen kirjasi käyttämänsä hakulausekkeet ylös taulukoihin. Tutkimustapa vaihteli riippuen osumien määrästä, osuvuudesta tai ajankohdasta. Useimmiten hakuprosessi oli luontevaa aloittaa artikkeleista, joissa oli kuvattu kohteen historiaa. Niiden pohjalta saatiin tietoa kohteen luonteesta ja tapahtumien ajankohdista. Tämän jälkeen kohteen kronologinen tutkiminen oli mahdollista, koska soveltuvia aineistoja voitiin hakea kohteen alkamisajankohdan tienoilta ja lähteä käymään läpi aineistoja nykypäivää kohti.

Myöhemmistä artikkeleista saatiin uusia hakuideoita koskien kiinnostavia tarinanaihoita, joita voitiin hakea jälleen aiemmista julkaisuista. Aineistojen louhintaa tapahtui siis sekä kronologisesti että retrospektiivisesti. Näin saatiin kattava kuva yritysten historiasta ja niihin liittyvistä digitaalisista aineistoista. Yksittäisistä yrityksistä tehtiin vähintään kymmenen hakua, ennen kuin merkittäviä tarinanaihoita löytyi.

Yksittäiset leikkeet valikoitiin sadoista osumista sen perusteella, kuinka arvioimme niiden kiinnostavuutta yritysten näkökulmasta. Esiin nostetut tarinanaihiot ovat kuvauksia, jotka liittyvät 1) henkilöihin 2) paikkoihin 3) tapahtumiin tai toimintoihin, sekä tietyn edellytyksin myös 4) aikaan ja 5) visuaalisuuteen. Suurin osa aineistosta oli julkaistu Etelä-Savossa, mikä korosti näkökulmien paikallisuutta. Isompien uutisten osalta aineistoa löytyi myös maakunnan ulkopuolisista julkaisuista. Kaikkiaan leikkeitä käytiin edellä kuvatulla tavalla läpi noin 750 kpl, joista noin 300 kpl oli ennen vuotta 1930 julkaistuja. Näitä tarinanaihion sisältäviä leikkeitä sekä tutkimuksen sivutuotteena kertynyttä tietämystä yritysten aikajanasta hyödynnettiin myöhemmin yritysten kanssa käydyissä keskusteluissa sekä jatkoideoiden yhteiskehittelyssä.

## YRITYSTEN KANSSA TOTEUTETTU YHTEISKEHITTÄMINEN

Valikoimme edellä kuvattujen hakujen perusteella 77 potentiaalista yritystä ja yhteisöä tarkempaan tarkasteluun. Ne edustivat melko kattavasti eri toimialoja ja alueita Etelä-Savossa (taulukko 1). Seuraavassa vaiheessa otimme lähempään tarkasteluun mukaan vain yrityksiä.

*Taulukko 1. Yritys- ja yhteisökohteiden jakautuminen toimialoittain (Tilastokeskuksen toimialaluokitus 2008)*

Toimiala	Kohteita kpl (n = 77):
Majoitus- ja ravitsemistoiminta	25
Muu palvelutoiminta	14
Teollisuus	7
Maatalous, metsätalous ja kalatalous	6
Taiteet, viihde ja virkistys	5
Tukku- ja vähittäiskauppa; moottoriajoneuvojen ja moottoripyörien korjaus	4
Kuljetus ja varastointi	3
Toimiala tuntematon	3
Julkinen hallinto ja maanpuolustus; pakollinen sosiaalivakuutus	3
Sähkö-, kaasu- ja lämpöhuolto, jäähdytysliiketoiminta	1
Informaatio ja viestintä	1
Rahoitus- ja vakuutustoiminta	1
Kiinteistöalan toiminta	1
Ammatillinen, tieteellinen ja tekninen toiminta	1
Koulutus	1
Terveys- ja sosiaalipalvelut	1

Löysimme riittävästi digitaalista aineistoa erityisesti yrityksistä, jotka kuuluivat majoitus- ja ravitsemistoiminnan (25 kpl) toimialaan. Monet niistä olivat kartanoita ja muita kulttuurihistoriallisesti merkittäviä rakennuksia, jotka on mukautettu palvelemaan majoitus- ja/tai ravitsemistoimintoja. Näihin yrityksiin liittyvät tarinat ovat luonteva osa yrityskuvaa ja siten digitaalisista aineistoista löytyvät tarinanaihiot sopivat hyvin niiden markkinointiin. Useilla yrityksillä oli myös selkeä paikkasidos suhteessa löydettyihin tarina-aihioihin.

Potentiaalisilla yhteistyöyrityksillä oli pitkä historiallinen perinne, josta louhimme esiin arviomme perusteella mielenkiintoisia löytöjä. Tarvittiin vielä yritysten oma halukkuus hyödyntää niitä. Otimme yhteyttä yrityksiin pääosin sähköpostitse. Useimmissa tapauksissa yritysten

edustajat suhtautuivat asiaan kiinnostuneesti. Tyypillisesti saimme luotua alkukontaktin suoraan yritysten omistajiin, jotka olivat mukana keskusteluissa muun markkinoinnista vastaavan henkilöstön kanssa. Yhteistyön konkretisoituminen ensikontaktista alkutapaamiseen vei kuitenkin aikaa. Hyödynsimme tämän ajan etsimällä lisää mahdollisimman kohdennettua aineistoa kohteen tarinoista, jotta yhteistyö yrityksen kanssa voitaisiin aloittaa tehokkaasti.

Yrityskäynnit (13 kpl) toteutettiin yhtenäisen konseptin mukaisesti. Aluksi tiedustelimme yritysten edustajilta, tunsivatko he Kansalliskirjaston Digi.kansalliskirjasto.fi -sivustoa entuudestaan. Yhtä lukuun ottamatta digitaaliset aineistot tai niiden hyödyntäminen eivät olleet tuttuja. Tämän jälkeen esittelimme sivustoa ja nostimme samalla esiin löytämiämme, yritykseen liittyviä tarinanaihiolkeikkaita. Yritykset odottivat tapaamisissa esimerkkejä löydöksistä. Kun yritysten edustajat olivat ymmärtäneet sivuston toimintaperiaatteen ja löydetyn aineiston mahdollisuudet, keskustelut siirtyivät yritystä koskevien digitaalisten aineistojen hyödyntämiseen. Vastaanotto yrityksissä oli poikkeuksetta kiinnostunutta, jopa innostunutta. Leikkeitä selailtiin yleensä hyväntuulisesti ja joskus uutiset saivat huumorin kukkimaan. Yritykset olivat pääsääntöisesti hyvin kiinnostuneita niitä koskevien löydösten hyödyntämisestä. Yrityskäynneillä käytyjen keskustelun ja noissa tilanteissa tekemiemme havaintojen perusteella voimme sanoa, että historialliset digitaaliset lehtiaineistot tarjoavat uusia, jopa yllättäviä näkökulmia yrityksille. Kyse on resurssista, jota voidaan hyödyntää paitsi yritysten brändien rakentamiseen ja uusien tuotteiden kehittelyyn, myös yritysten identiteettien vahvistamiseen ja jo unohtuneen tiedon hankkimiseen yrityksestä.

Keskeistä yhteiskehittelyprosessissa on jännite, joka muodostuu historiallisten ajankuvien tarjoamien näkökulmien ja nykyään tuntemiemme yritysten ja paikkojen todellisuuksien välille. Kuvaukset, jotka heijastelevat aikoinaan eläneiden ihmisten kokemuksia ja näkemyksiä, poikkeavat usein yllättävällä tavalla omasta kokemusmaailmastamme. Tämä jännite loi tapaamisissa hedelmällisen ilmapiirin kehitellä ja kuvitella uusia tapoja käyttää yrityksen menneisyydestä nousevia tarinoita ja valokuvia. Yritystapaamisissa keskustelu saattoi ohjautua täysin odottamattomaan suuntaan hyvinkin pienillä huomioilla tai löydöksillä, jolloin ilmapiiri vapautui ja innovaatioaihiota alkoi muodostua (kuva 3).



Kuva 3. Mikkelin Osuuskaupan avajaisista kertova mainos vuodelta 1916 (Mikkelin Sanomat, 03.08.1916, nro 84, s. 1)

Tekijänoikeuksiin liittyvät epäselvyydet, kuten lehtiaineistojen jakautuminen avoimiin ja rajattuihin aineistoihin sekä valokuvien käytettävyyteen liittyvät tekijänoikeuskysymykset, vaikeuttavat käytännössä aineiston sujuvaa hyödyntämistä. Tämä ongelma on tullut esiin jo aiemmin, kun on etsitty ratkaisuja avoimen datan hyödyntämiseen yritysten liiketoiminnassa (Rajahonka & Villman 2018, 32–33). Yritykset ovat kyllä valmiita maksamaan kohtuullisena pidetyt korvaukset käyttökelpoisista aineistoista. Oma haasteensa yrityksissä liittyy ajankäyttöön. Alkutapaamisten jälkeen yrittäjillä tai niiden edustajilla oli vain rajatusti aikaa jatkaa itsenäisesti aineiston läpikäyntiä ja hyödyntämisen pohtimista.

## **TULOSENA TARINA- JA INNOVAATIOAIHIOITA MENESTYVILLE YRITYKSILLE**

”Storytelling is a combination of retrospective and prospective sensemaking.”

(Boje 2014, Preface xxix)

Kaikkia yritystapaamisia yhdistänyt havaintomme on, että *tarinoiden merkitys tunnistetaan yrityksen markkinoinnin sekä tuotteiden ja palvelujen kehittämisen tärkeänä osana*. Tarinoiden kerronta (storytelling) nähdään nykyisin oleellisena osana menestyvän yrityksen brändiä ja liiketoimintaa (Fog ym. 2010).



Historiallisista sanoma- ja aikakauslehtiaineistoista esiin nostamamme esimerkit kytkeytyivät yritysten kanssa käydyissä keskusteluissa luontevasti niiden näkemykseen omasta identiteetistään, vahvuuksistaan ja erityispiirteistään. Yritykset olivat oivaltaneet nämä tekijät tärkeäksi osaksi omaa kilpailukykyään. Esimerkiksi matkailuyritys voi etsiä erityisyyttä paikallisesta toimintaympäristöstään halutessaan erottua muista alan yrityksistä. Paikan historia, tapahtumat, kulttuuri ja ihmiset ovat tärkeitä yritysten brändin ja tarinan kerronnan osia.

”Hyvä brändi on kuin tarina, joka koskettaa, naurattaa, herättää tunteita ja on ihmisseläinen. Sanotaan, että hyvän tarinan muistaa toisin kuin yksittäistä tilastollista faktaa. Hyvää tarinaa myös kerrotaan mielellään eteenpäin, jolloin yrityksen asiakkaista ja muista sidosryhmistä tulee aktiivisia brändimielikuvan rakentajia.” (Törmälä 2018)

Tarinoita voidaan hyödyntää monilla eri tavoilla. Ravintola voi yhdistää paikallisuutta painottavaan menuunsa tarinoita. Sähköyhtiö voi nostaa esiin omia juuriaan paikallisena, sosiaalista vastuuta pitkään kantaneena yrityksenä, ja linja-auto- tai laivayhtiö voivat nostaa uusina tuotteina esiin historiallisia reittejä, joita vintage-tyyliin paketoidaan historiaharrastajille. Matkailukohde voi löytää lähistöstään mielenkiintoisia paikkoja, joihin liittyy jännittäviä historiallisia tapahtumia, ja joita se voi hyödyntää esimerkiksi matkailutuotteissaan. Vanhat valokuvat voivat tarjota pihasuunnitteluun uutta tietoa menneestä ympäristöstä.

Jokaisella yrityksellä on oma erityinen historiansa, jota ei voi kopioida. Tässä esitelty historiallinen aineisto auttaa yrityksiä syventämään tätä erityistä osaa identiteetistään ja rikastamaan sitä uusilla näkökulmilla. Aineistossamme oli monia pitkän historian omaavia perheyrityksiä, jotka ovat hyvin tietoisia historiastaan. Yrityshistoriaa oli myös selvitetty, kirjattu ja säilötty kuvina omiin arkistoihin. Tästä huolimatta esittelemämme digitaalinen aineisto toi esiin uusia piirteitä yritysten omasta historiasta. Aineisto mahdollistaa moniäänisen tarinankerronnan, vähintään uuden näkökulman johonkin jo tunnettuun yrityksen historian vaiheeseen.

Journalistinen dokumentointi historian tapahtumista herätti siis suurta mielenkiintoa yrityksissä. Kansalliskirjaston avointa verkkopalvelua keuhuttiin hyödylliseksi välineeksi jatkaa oman historian tutkimista. Yhteistyö yritysten kanssa jatkuu edelleen. Vireillä on myös muutamia konkreettisia lehtiaineistoa hyödyntäviä innovaatioita, joita yritykset ovat päättäneet toteuttaa.

Kansalliskirjaston digitaalinen historiallisten aineistojen käytön verkkosivu (<https://digi.kansalliskirjasto.fi>) näyttää avaavan kansalaisille ja yrityksille aarretaiton omien identiteettiresurssien kehittämiseen. Sivustoa ei kuitenkaan tunneta. Olisi hyvin tärkeätä, että tämä mahdollisuus tiedostettaisiin nykyistä laajemmin, avoin käyttöliittymä tulisi entistä useamman yritysten ja kansalaisten käyttöön, ja tekijänoikeuskysymykset voitaisiin ratkaista eri osapuolia hyödyttävällä tavalla.

## LÄHTEET

Alasuutari, P. 2017. Tasavalta - Sodan jälkeisen Suomen kaudet ja trendit. Tampere: Vastapaino.

Boje, D.M. 2014. *Storytelling Organizational Practices. Managing in the quantum age.* New York: Routledge.

Fog, K., Budtz, C., Munch, P. & Blanchette, S. (Eds.) 2010. *Storytelling. Branding in Practice.* Copenhagen: Springer.

Kitch, C. 2008. Placing journalism inside memory – and memory studies. *Memory Studies* 2008, 1: 311.

Rajahonka M. & K. Villman. 2018. Caseja ja tietoa avoimesta datasta: DigiBus – Liiketoimintamahdollisuuksia digitaalisista aineistoista. *Xamk kehittää* 38, Kaakkois-Suomen ammattikorkeakoulu Oy. Saatavilla: <http://urn.fi/URN:ISBN:978-952-344-078-4>.

Seuri, O. 2014. Historian jokapäiväinen virta sanomalehdessä. *Media & viestintä* 37, 3, 22–37.

Törmälä, M. 2018. Brändi ei ole logo – se on hyvä tarina. Oulu Business School. Oulun yliopisto. Saatavilla: <https://www oulu.fi/blogs/node/42508>

## LEIKEKUVIEN LÄHTEET

Suomen Kuvalehti, 21.03.1925, nro 12, s. 13

<https://digi.kansalliskirjasto.fi/aikakausi/binding/889576/articles/3501744?page=13>

Kansalliskirjaston digitaaliset aineistot

Vapaus, 07.11.1908, nro 125, s. 4

<https://digi.kansalliskirjasto.fi/sanomalehti/binding/811384/articles/3211608?page=4>

Kansalliskirjaston digitaaliset aineistot

Mikkelin Sanomat, 03.08.1916, nro 84, s. 1

<https://digi.kansalliskirjasto.fi/sanomalehti/binding/1271631/articles/2959631?page=1>

Kansalliskirjaston digitaaliset aineistot

# TIETO HYÖTYKÄYTTÖÖN TIEDOLLA JOHTAMISTA KEHITTÄMÄLLÄ

Miia Kosonen, KTT, tki-asiantuntija, Xamk

Tieto saa arvonsa käytössä. Näistä lähtökohdista Digitalia teki Digitaaliset aineistot käyttöön -hankkeessa yhteistyötä alueen yritysten kanssa tiedolla johtamisen valmiuksien parantamiseksi.

Viimeistään nyt organisaatioiden on aika sisäistää, että aineeton pääoma on aikamme kovinta valuuttaa. Professori Pirjo Ståhlen mukaan hyvät tietojohdajat huomioivat ihmisten, teknologian ja liiketoiminnan muodostaman kokonaisuuden, yhdistäen teknologisen ja sosiaalisen arkkitehtuurin saumattomasti yhteen (Kosonen 2019a). Aidosti monitieteisenä alana tietojohdaminen pitää sisällään kaikki ne käytännöt, joilla parannetaan edellytyksiä luoda arvoa aineettomista resursseista, oli kyseessä sitten inhimillinen tai sosiaalinen pääoma, rakenteellinen data tai big data, historialliset aineistot tai nykyhetken tieto. Kaikki Digitaliassa tehtävä työ kytkeytyy siten hyvin vahvasti tietojohdamiseen.

## TIEDOLLA JOHTAMINEN

Yksi julkisessa keskustelussa näkyvimmistä tietojohdamisen osa-alueista on *tiedolla johtaminen* eli analysoidun tiedon hyödyntäminen liiketoiminnassa ja päätöksenteossa (Kosonen 2015). Tiedolla on aina oma konteksti, jota on välttämätöntä ymmärtää ja tulkita – siksi yleispätevää kaikille sopivaa patenttiratkaisua parempaan johtamiseen ei ole olemassa. Peruskäsitteiden selkiyttämiseksi hankkeessa on kuitenkin tuotettu avoin ja yleistajuinen tiedolla johtamisen opas niin pk-yrityksille kuin muillekin aiheesta kiinnostuneille (Kosonen 2019b).

Tiedon tuottaminen ja säilöminen on vasta kolikon toinen puoli. Yksilötasolla tarvitaan motivaatio, mahdollisuudet ja kyky hyödyntää tietoa, organisaation ja yhteiskunnan tasolla tieto-orientoitunut kulttuuri. Käytäntö on harvoin näin ihanteellinen ja arjessa joudutaan myös sovittamaan yhteen ristiriitaisia tavoitteita ja intressejä. Pääkkönen ja Lilja (2019) sekä Hyyryläinen ym. (2019) mainitsevat tässä artikkelikokoelmassa esimerkkinä tekijänoikeuksien asettamat rajoitteet historiallisten aineistojen jatkokäytölle yrityksissä.

Miksi tiedon tuottamisen ja hyödyntämisen silmukka epäonnistuu usein? Jalosen (2015) mukaan informaatiota ylituotetaan systemaattisesti ja sitä vältetään tietoisesti, minkä lisäksi organisaatioissa pyritään välttämään moitteita. Jos pystyisimme tekemään näkyväksi kaikki ne kustannukset, joita esimerkiksi informaation ylituotanto aiheuttaa, saattaisi turha raportointi ja tiedon hautaaminen erilaisiin tietojärjestelmiin loppua saman tien.

Ongelmien ratkaiseminen tiedon pohjalta edellyttää käsitystä siitä, millaisesta ongelmasta on kyse. Tieto-ongelmat voidaan jakaa neljään eri tyyppiin: epävarmuus, monimutkaisuus, epäselvyys ja monitulkintaisuus. Usein organisaatioissa pidetään haasteena tiedon puuttumista, ts. epävarmuutta, vaikka tietoa olisi jo saatavilla vaikka kuinka. Sitä vain ei osata, haluta tai kyetä hyödyntämään (Feldman & March 1981).

Davenport ja Harris (2007) kuvaavat viisiportaisen mallin tiedon hallintaan. Davenport on esittänyt, että valtaosa yrityksistä on tasolla kaksi. Ne siis keräävät satunnaisesti tietoa tietämättä itsekään syytä siihen.

- Mihin ihmeessä me keräisimme tietoa
- Yritys kerää satunnaisesti tietoa tietämättä miksi
- Yrityksellä on tehokkaat tiedonkeruujärjestelmät, mutta se ei hyödynnä keräämäänsä
- Yritys kerää tietoa systemaattisesti ja hyödyntää siitä osan
- Yritys kilpailee tiedon avulla

Tietojohdetun organisaation rakentamisessa yksi avainkysymys onkin oikeiden metriikoiden valinta. Esimerkiksi nk. *vanity metrics* on uskallettava heittää menemään. Organisaatiot muodostuvat ihmisistä, ja ihmisten tavoin niillä on taipumusta turhamaisuuteen: saatetaan mitata esimerkiksi sovellusten latauskertoja ja artikkeleiden klikkauskäyttöä. Hedelmällisempi lähtökohta on pyrkiä ennakoimaan tulevaa ja tähdätä sen kautta kasvuun (kts. Laihonon ym. 2015).

## **YRITYSYHTEISTYÖ DIGITAALISET AINEISTOT KÄYTTÖÖN -HANKKEESSA**

Tiedolla johtamisen kokonaisuus käynnistettiin kartoittamalla yritysten kehittämistarpeet lyhyellä verkkokyselyllä ja face-to-face -keskusteluissa. Matkailuyrityksille toteutettiin marraskuussa 2018 kolme digivalmennusta, kukin kestoltaan 2,5 h. Osallistujia oli kaikkiaan 6 henkeä. Valmennuksessa asialistalla olivat verkkosivut, sosiaalisen median hyödyntäminen ja analytiikka. Kokonaisuudesta pyrittiin erottelamaan olennainen tieto asiakkaiden ja oman liiketoiminnan kehittämisen kannalta, rajaamaan vain siihen ja karsimaan rönsyt. Osallistuvat yritykset olivat Lakeland GTE, Anttolanhovi ja Löydön Kartano.

Työpajat koettiin erittäin hyödyllisinä ja konkreettisina. Erään osallistujan mukaan

*“...pienyrittäjällä on harvoin mahdollisuus saada kokonaisvaltaista palautetta, tämä oli ensimmäinen kerta, ja sain paljon kehittämisideoita. Paino sanalla kokonaisvaltainen! Nykyään tietoa tulvii joka suunnasta ja se on pirstaleista. Oli hyvä istua hetki sivuston äärellä ja käydä kokonaisuutta läpi. Sain konkreettisia ehdotuksia joita voin nyt työstää. Mielestäni jokaisen yrittäjän pitäisi käydä läpi vastaava sparraus.”*

Edellä oli esillä tiedon kontekstuaalisuus. Siksi tiedolla johtamista on haastavaa ellei mahdollonta edistää ilman, että istutaan yhteisen pöydän ääreen tekemään ja kokeilemaan – se on toki hitaampaa, mutta lopputulos on parempi kuin pintapuolisia oppeja ulkoapäin kaatamalla. Tämä näkyi myös matkailuyrityksiltä saadussa palautteessa.

OP Suur-Savon osalta ratkaisuna olivat Xamkin ja pankin yhteiset Digimessut, käytännönläheiset valmennukset henkilöstölle ja hallinnolle. Digimessujen sosiaalisen median osuudessa oli mukana yli 70 osallistujaa neljällä eri paikkakunnalla, Mikkeli, Savonlinna, Juva ja Pieksämäki. Työpajoja järjestettiin kaikkiaan 9 (yhden työpajan kesto 2,5 h) tammi-helmikuussa 2018 ja toukokuussa 2018.

### **CASE: OP SUUR-SAVON DIGIMESSUT**

Hankkeessa on tuotu OP Suur-Savon toimintaan mukaan perinteisten toimintomittarien (näytöt, jaot, kommentit ym.) lisäksi mittarit myös sille, kuinka suuri osa henkilöstöstä on aktiivinen sosiaalisen median kautta tehtävässä myynnissä ja asiakaspalvelutyössä, ja osaa hyödyntää perusanalytiikkaa oman työnsä kehittämiseksi. Tämä ei ole pelkästään markkinointi- ja viestintätiimin asia, kuten usein ajatellaan, vaan tiedon ja asiantuntemuksen jakamisen on oltava koko organisaation yhteinen tavoite – ja erottamaton osa tietokulttuuria. Alkukartoituksessa tunnistetuista kehittämistarpeista otettiin lähempään käsittelyyn analytiikan hyödyntäminen sekä henkilöstön aktivoiminen sosiaalisen median kanaviin.

Entä tulokset Digimessuilta? Osalla työpajojen osallistujista aktiivisuus lisääntyi ja on myös pysynyt yllä. Palautteen mukaan ihmiset saivat teknisen osaamisen lisäksi “henkistä pääomaa ja rohkeutta”. Viestintä- ja markkinointitiimi havaitsi, että Facebook-sivujen postauksia jaetaan useammin. Tässä myös uuden toimitusjohtajan aktiivisuus on toiminut esimerkkinä. Sosiaalisen median kanavat ovat kokonaisuutena tulleet luontevammaksi osaksi erilaisia keskusteluja ja pankin arkea. Mikä tärkeintä, julkaisemisen ja jakamisen lisäksi monet ovat ryhtyneet seuraamaan erilaisia ammattiin liittyviä someryhmiä tai keskusteluvustoja kehittämään osaamistaan. Eniten muutosta tapahtui tiimin mukaan toimihenkilöiden kohdalla, hallinnossa ja myynninjohdossa vähemmän.

Varainhoitaja Essi Sarkakari-Kosamo kertoo kokemuksistaan ja oppimastaan:

*Olen Essi Sarkakari-Kosamo (@EssiSarkakari) ja työskentelen varainhoitajana OP Suur-Savon Private-tiimissä. Hoidan pääasiassa henkilöasiakkaiden sijoitusvarallisuutta. Olen ollut aktiivinen sosiaalisessa mediassa jo useamman vuoden sijoitusaiheisissa keskusteluissa, erityisesti Facebookin sijoittamiseen liittyvissä ryhmissä ja LinkedInissä. Kokeneena somettajana olin digityöpajaan tullessani hieman ennakkoluuloinen. Miia kuitenkin rohkaisi kokeilemaan uutta ja siitäpä se sitten lähti. Digipajan jälkeen olen esimerkiksi saanut noin 900 Twitter-seuraajaa lisää ja siitä on tullut yllättäen keskeisin*

*sijoitusaiheinen kanavani. Taloudellinen hyvinvointi kuuluu jokaiselle ja sitä sanomaa haluan kertoa eteenpäin. Samalla olen tutustunut erilaisiin sijoittajiin ja sijoitustyyliihin, ja se on mahtavaa. Olen myös saanut pari vinkkausta mielenkiintoisista työpaikoista ja jopa työtarjouksia. Parhaimmillaan twiiteilläni on ollut näyttökertoja kuukauden aikana yli 100 000, edellisessä kuussa noin 50 000. Kertaalleen olen myös noussut suosittujen sijoitustwiittajien top 10 -listalle. Sosiaalisessa mediassa keskustelemalla ymmärtää paljon paremmin asiakkaiden todelliset tarpeet, tarpeiden muutokset ja tyytyväisyyteen vaikuttavat tekijät, mitä he eivät välttämättä halua sanoa suoraan pankissa käydessään. Parhaat jutut löytyvät ihan tavallisesta keskustelusta muiden sijoittajien kanssa ja aktiivisesti kommentoiden.*

OP Suur-Savon mukaan Digitalia-yhteistyössä parasta oli sen vaivattomuus ja räätälöitävyys juuri pankin tarpeisiin. Kuitenkin sisältö oli sellaista, jota mikä tahansa yritys voisi hyödyntää omalta näkökannaltaan ja omiin tavoitteisiinsa peilaten. Toteutukseen oli varattu hyvin aikaa, suunnittelu aloitettiin ajoissa ja se sujui mutkattomasti. Materiaalit ja selkeät ratkaisuehdotukset tulivat Xamkilta. OP Suur-Savolle oli tärkeää myös yhteistyön paikallisuus, koska molemmat ovat eteläsavolaisia toimijoita.

Osallistujilta kerättiin pankin intranetin kautta palautetta ja työpajat saivat vastaajilta paljon kiitosta. Erityisesti mainittiin vinkit, joilla omaa työtä ja asiantuntemusta voi lähteä tuomaan esille. Kehitettävää -osastolla epäiltiin, että edistyneempien työpaja meni jo liian syvälle. Kaikkien maailmaan ei vielä mahdu ajatus näkymisestä ja vaikuttamisesta verkossa myös ammatillisessa mielessä. Työtä riittää siis tulevaisuudessakin.

Yhteistyöstä viestittiin sekä pankin että Xamkin sähköisissä kanavissa (kts. Kosonen & Wase-nius 2018). Lähitulevaisuudessa yrityksen tavoitteena on viedä eteenpäin omaa videotuotantoa ja laatia sosiaalisen median ohjeistuksia. Aina ajankohtaisia teemoja ovat henkilöstön osallistaminen ja paikallisuuden vahvempi korostaminen. Tämä näkyy alueen eri toimijoiden kanssa tehtävän yhteistyön laajuudessa ja syvyydessä.

## **LOPUKSI**

OP Suur-Savossa tavoitteeksi asetettiin saada pankin henkilöstö ja hallinto käyttämään sosiaalisen median kanavia nykyistä aktiivisemmin. Näin tieto leviäisi tehokkaammin ja vaikutukset ulottuisivat laajemmalle paikalliseen yhteisöön. Tässä onnistuttiin hyvin. Tiedolla johtaminen ei ole pelkästään näkyvän tiedon johtamista, vaan tärkeässä roolissa on myös hiljainen, kokemuksellinen tieto. Tämä näkyi OP Suur-Savossa oivallisesti: pankin henkilöstö on lähtenyt kehittämään asiantuntemustaan liittymällä erilaisiin sosiaalisen median ryhmiin ja aiheyhteisöihin yli yritysrajojen. Kuten varainhoitajan esimerkki edellä havainnollisti, keskustelujen kautta aukeaa ikkuna asiakkaiden tarpeiden syvempään ymmärtämiseen ja myös tulevaisuuden tarpeiden ennakointiin.

Tiedolla johtamisen kehittämistä voidaan organisaatioissa toteuttaa kolmella tasolla. Ensimmäisellä eli **perustasolla** organisaatiolla ei vielä ole näkemystä siitä, mitä ylipäättään on tiedolla johtaminen. Esimerkkinä, jos sosiaalisen median kanavien tuntemusta on vähän, organisaatiolta useimmiten puuttuu sosiaalisen median sisältö, ja jos ei ole sisältöä, ei tarvita myöskään analytiikkaa ja tiedon vaikuttavuuden arviointia. Toisella eli **analytiikkatasolla** ilmiö on jossain määrin tuttu, mutta konkreettisten työvälineiden kanssa saatetaan tarvita lisäoppia. Tätä tasoa hankkeessa edustivat muun muassa myynninjohtajan työpajat, joissa käytiin läpi tavoiteasetantaa ja seuranta sosiaalisen median kanavissa.

Kolmannella tasolla – kutsuttakoon sitä vaikkapa **data scientist -tasoksi** – mukaan tulevat digitaalisen datan hyödyntämisen laskennalliset menetelmät ja tutkimuskäyttö. Kuvailevan ja tietyn organisaation omista lähtökohdista tehtävän analytiikan sijaan pyritään ymmärtämään laajempia kokonaisuuksia ja ennakoimaan tiedon avulla toimintaympäristön muutoksia (kts. Isotalo ym. 2019). Data ei analysoi itse itseään eikä vastauksia kysymyksiin edelleenkaan saada nappia painamalla: tarvitaan paitsi ihmisten ja koneiden yhteistyötä myös ymmärrystä niistä alustoista, joista data on lähtöisin. Yrityksillä ei välttämättä ole resursseja toteuttaa kaikkia kolmea tasoa, mutta trendien haistelussa ovat apuna verkostot ja yhteiset projektit esimerkiksi tutkimusorganisaatioiden kanssa.

Lähivuosina tiedolla johtamisen kehittäminen Etelä-Savon alueella jatkuu muun muassa Digiportaati-hankekokonaisuudessa 2019-2020. Digiportaati tarjoaa maakunnan yrityksille tarvelähtöistä digikoulutusta Mikkelissä ja Savonlinnassa. Ihmisillä ja tiedolla on paljon yhteistä: kumpikaan ei tule koskaan valmiiksi, ja molemmissa tapauksissa johtaminen tarjoaa kiinnostavia haasteita vuosikymmeniksi eteenpäin.

## LÄHTEET

- Davenport, T. & Harris, J. 2007. *Competing on Analytics: The New Science of Winning*. Boston: Harvard Business School Publishing.
- Feldman, M. & March, J. 1981. Information in organizations as signal and symbol. *Administrative Science Quarterly*, 26(2), 171-186, June 1981.
- Isotalo, V., Laaksonen, S-M., Pöyry, E. & Jallinoja, P. 2019. Sosiaalisen median ennustekyky kaupan myynnissä – esimerkkinä veganismi ja vegaaniset ruuat. *Kansantaloudellinen aikakauskirja* 115(1), 91-112.
- Jalonen, H. 2014. A framework for dealing with fundamental knowledge problems through social media. *VINE* 44(4), 558-578.
- Jalonen, H. 2015. Tiedolla johtamisen näyttämö ja kulissit. Teoksessa Virtanen, P. Stenvall, J. & Rannisto, P-H. (toim.), *Tiedolla johtaminen – teoriaa ja käytäntöjä*, pp. 40-68. Tampere: Tampere University Press.
- Kosonen, M. 2015. Tietojohdaminen ja tiedolla johtaminen. WWW-dokumentti. Saatavissa: <https://www.slideshare.net/miiak/tietojohdaminen-ja-tiedolla-johtaminen> [viitattu 25.4.2019].
- Kosonen, M. & Wasenius, T. 2018. Osuuspankkilaiset rohkeasti some-oppien äärellä. *Kaakkois-Suomen ammattikorkeakoulun Read-verkkolehti* 1/2018.
- Kosonen M. 2019a. Tietojohdamisen matka bisneksen kovaan ytimeen. *Tietoasiantuntija* 1/2019, 26-27.
- Kosonen M. 2019b. Tiedolla johtamisen käsikirja. WWW-dokumentti. Saatavissa: <https://digitalia.xamk.fi/tijo/> [viitattu 30.4.2019].
- Laihonen, H., Lönnqvist, A. & Metsälä, J. 2015. Two knowledge perspectives to growth management. *VINE*, 45(4), 473-494.



# DIGITALIA LUO MAHDOLLISUUKSIA

Noora Talsi, YTT, tutkimusjohtaja & Miia Kosonen, KTT, tki-asiantuntija,  
Kaakkois-Suomen ammattikorkeakoulu

Tässä artikkelikokoelmassa on luotu katsaus Digitaaliset aineistot käyttöön –hankkeen tuloksiin. Alkupuolen artikkelit käsittelevät digitaalisessa muodossa olevan tiedon käytettävyyden ja käyttäjäkokemuksen parantamista. Yksi hankkeen tärkeä tavoite oli laajentaa aineistojen käyttömahdollisuuksia eri käyttäjäryhmille. Niinpä artikkeleissa syvennyttiin myös digitointiin, aineistojen käyttöoikeuksiin sekä digitaalisen tiedon ja tarjolla olevien palvelujen hyödyntämiseen yrityksissä ja niiden liiketoiminnassa.

Digitaalisuus on monialainen ilmiökenttä, jossa monitieteisyys ei edelleenkään ole pelkästään toivottavaa vaan välttämätöntä. Tämä tavoite on Digitaliassa toteutunut hyvin. Vahvuksiamme ovat tieteenala- ja oppilaitosrajat rikkova monialainen yhteistyö ja laajat verkostot alueellisesti, kansallisesti ja kansainvälisesti. Näkemyksemme mukaan koko digitaalisuuskehityksen ydin on tiedossa ja mahdollisuuksissa hyödyntää sitä. Näitä edellytyksiä Digitalia haluaa parantaa jatkossakin.

Entä millaiset ovat Digitalian tulevaisuuden näkymät? Digitalia on vahvasti mukana Memory Campuksen kehittämisessä. Memory Campus on arkisto- ja kirjastotoimialan ja digitoinnin osaamis- ja palvelukeskus, jonka olennaisia osia ovat tutkimus, aineistojen käytettävyyden ja saatavuuden parantaminen, tuote- ja palveluliiketoiminnan kehittäminen ja alan korkeakoulutus. Kehittämisalustan omistaa Mikkelin kaupunki ja Miksei koordinoi yhteistyötä. Xamkin, Helsingin yliopiston ja Kansalliskirjaston osaaminen ja kehitystyö tukevat hyvin Memory Campuksen kokonaisuutta.

Kansallisarkiston kanssa tehtävää yhteistyötä on Digitaaliset aineistot käyttöön -hankkeessa vahvistettu liittyen massadigitoinnin mahdollistamiseen. Valtioneuvosto on tehnyt periaatepäätöksen asiakirjallisen aineiston digitoinnista ja arkistoinnista vain sähköisenä. Nyt Kansallisarkisto tähtää siihen, että arkistoitava analoginen aineisto digitoidaan Mikkelissä. Digitointi avaa aivan uusia mahdollisuuksia aineiston käyttöön ja hallintaan, mutta asettaa vaatimuksia tehokkaiden ja automaattisten arkistointiratkaisujen kehittämiseksi. Yhteistyötä on jo tehty liittyen asiakirjojen PDF-pilkontaan ja sitä jatketaan edelleen (Jääskeläinen ja Räisänen 2019, tässä kokoelmassa). Digitaliassa tehtävä työ tukee paitsi Kansallisarkistossa tehtävää massadigitointia, myös muita tahoja, jotka ovat kiinnostuneita digitaalisen tiedon hyödyntämisestä ja sen käytettävyyden parantamisesta.

Kansalliskirjaston digitoitu aineisto toimii tulevaisuudessa tutkimuksen lähteenä entistä laajemmin Suomessa ja mahdollisesti myös kansainvälisesti. Aineiston kansalaiskäyttö lisääntyy edelleen ja aineistoa käyttävät myös uudet käyttäjäryhmät. Kansalliskirjasto tähtää jatkossa library labin eli digitaalisten aineistojen käyttäjä- ja kehittäjäyhteisön luomiseen. Luomiseen tarvitaan aineistojen hyödyntäjien käyttäjätutkimusta tiedon käyttötapojen ja -tarpeiden selvittämiseksi, jotta saadaan malli käyttäjäryhmien analysoinnille, vuorovaikutukselle ja yhteistyölle. Digi.kansalliskirjasto.fi:n eli Kansalliskirjaston digitaalisen kirjaston kehitystyö jatkuu myös edelleen.

Yksi kehittämistyön kohteista Digitaliassa on digitaalisten aineistojen visualisointi, mistä esimerkkinä tässä artikkelikokoelmassa esitelty Helsingin pommitusten 3D-toteutus (Jääskeläinen 2019). Jatkossa Digitalia suunnittelee tuottavansa myös yleiskäyttöisemmän ratkaisun erilaisten virtuaalinäyttelyiden toteuttamiseen. Paikallisille oppilaitoksille ja yrityksille visualisoinnin ympärillä tehtävä kehittäminen voi avata mahdollisuuksia kiinnostaviin jatkopilotteihin esimerkiksi Etelä-Savon alueen matkailukohteita silmälläpitäen. Xamk on jo aiemmin tehnyt yhteistyötä museosektorin kanssa kolmiulotteisen aineiston digitoinnissa ja käsittelyssä, mistä esimerkkinä Mikkelin museoiden käytössä oleva ja verkosta löytyvä marsalkka Mannerheimin salonkivaunun 3D-malli (salonkivaunu.mikkeli.fi) sekä Viva3-projektissa kehitetty Astuvansalmen kalliomaalausalueen 3D-mallinnus.

Vaikka tekniset toteutukset ovat tärkeitä, niistä ei kannata tehdä itsetarkoitusta ja unohtaa ihmisiä: pelkkä asioiden siirtäminen haljuun 3D-ympäristöön – esimerkiksi tyhjän lentokoneen matkustamoon tai rakennukseen – tuskin vielä riittää kiinnostuksen herättämiseen ja kävijäryntäyksen luomiseen. Kaksi tärkeää näkökohtaa käyttäjien houkuttelemiseksi virtuaalitodellisuuteen ovat ns. mahdottomien asioiden tekeminen ja aikamatkailu, toisin sanoen saman fyysisen tilan esittäminen eri aikakausina. Kolmas ajuri on tuttujen ihmisten tuominen lähelle, mikä useimmista virtuaaliympäristöistä vielä toistaiseksi puuttuu. Käytännössä ihmiset edelleen matkustavat paikasta toiseen kokeakseen paikallista kulttuuria, tutustuakseen ihmisiin ja saadakseen makuelämyksiä. Tutkijoiden ja kehittäjien on tulevaisuudessa luotava mahdollisuuksia kohdata vaikkapa yhteisellä piknikillä Pariisissa tai Mikkelin torilla kavereiden kanssa 3D-teknologiaa hyödyntäen. Näistä toteutuksista myös ilmasto kiittäisi.

Lopuksi on vielä syytä mainita yhteistyö opetuksen ja koulutuksen kanssa. Digitalia on järjestänyt perinteiset tieteenalarajat rikkovaa kesäkoulua vuodesta 2016 alkaen. Kesäkoulun ideana on ollut tarjota digitaalisen tiedonhallinnan ja arkistoinnin opiskelijoille, tutkijoille ja arkisto- ja kirjastoalan ammattilaisille avoin asiantuntija- ja keskustelufoorumi.

Kesäkoulussa on käsitelty kysymyksiä, jotka liittyvät esimerkiksi digitaalisen tiedon eettisyyteen, juridisiin haasteisiin, tiedon luotettavuuden arviointiin ja sosiaalisessa mediassa liikkuvaan tietoon. Palaute tapahtumasta on ollut erinomaista ja jatkoa on toivottu. Kentällä on selkeästi tarve vapaamuotoiselle foorumille, jossa voi asiantuntijoiden opastuksella keskustella ajankohtaisista teemoista ja toteuttaa tiedonvaihtoa vertaisryhmissä. Keskustelu digitaalisen tiedon ympärillä jatkuu tulevissa kesäkouluissa – ja Digitalian tutkimushankkeissa.



