

MUISTIORGANISAATIOIDEN AINEISTON
TRANSKRIPTION JA SISÄLLÖNKUVAILUN
JOUKKOISTAMINEN

Marianne Ekman

Tampereen yliopisto
Viestintätieteiden tiedekunta
Informaatiotutkimus ja interak-
tiivinen media
Pro gradu -tutkielma
Toukokuu 2017

TAMPEREEN YLIOPISTO, Viestintätieteiden tiedekunta
Informaatiotutkimus ja interaktiivinen media
EKMAN, MARIANNE: Muistiorganisaatioiden aineiston transkription ja sisällönkuvailun joukkoistaminen
Pro gradu -tutkielma, 47 s., 1 liites.
Toukokuu 2017

Muistiorganisaatioiden aineistomäärä kasvaa koko ajan. Vähenevien resurssien kanssa kamppaileville muistiorganisaatioille ongelmaksi tulee aineiston saattaminen käyttökelteisessä muodossa yleisön käytettäväksi. Muistiorganisaatioissa on kokeiltu joukkoistamista eli suuren yleisön tehtäväksi annetaan internetin kautta ammattilaisten työtehtäviä.

Tutkielmassa selvitettiin, kuinka aineiston sisällönkuvailuun ja transkriptioon liittyviä joukkoistamishankkeita oli organisoitu ja haettiin organisointiin hyviä käytänteitä. Viitekehyksessä joukkoistaminen asetettiin arkistojen avautumisen ja vapaaehtoisten osallistamisen jatkumoon. Tutkimusaihetta tarkennettiin koskemaan käyttäjien rekrytointia, käyttäjien ja organisaation motivaatioita, hankkeen hallinnoijien ja käyttäjien rooleja sekä tehdyn työn laatua ja laadun kontrollia. Opinnäytteen metodina oli systemaattinen kirjallisuuskatsaus, jonka avulla löydettiin aineistoksi artikkelit tieteellisissä julkaisuissa vuosina 2012-2016.

Tutkimustulokset vahvistivat aiempia tutkimustuloksia. Motivoivana tekijänä käyttäjille oli oma kiinnostus aiheeseen, mutta myös halu osallistua yhteisen hyvän tekemiseen. Organisaatiot hyötyivät käyttäjien työpanoksesta. Koska käyttäjien osaamistaso ja tehtävien vaatimustaso vaihtelivat, niin joukkoistamisessa on kontrolloitava tehdyn työn laatua. Tutkielman tutkimustulokset osoittivat, että joukkoistamisessa pieni joukko teki suuren osan tehtävistä, joten oleellista olisi löytää nämä toimijat. Jos joukkoistaminen kohdentuu suurelle yleisölle, niin tehtävät olisi hyvä pilkkoa pieniksi osiksi. Asiantuntijoille kohdennettua joukkoistamista ja auktoriteetin jakoa voisi tutkia jatkossa lisää. Pelien käyttö tehtävän tekemisessä, laadun valvonnassa ja käyttäjien motivoinnissa nousi esiin mahdollisuutena, jossa on vielä tutkittavaa.

Avainsanat: muistiorganisaatiot, aineistot, joukkoistaminen, sisällönkuvailu, transkriptio,

Sisällys

1	JOHDANTO.....	1
2	TUTKIELMAN VIITEKEHYS	3
2.1	Joukkoistaminen ja arkistot	3
2.1.1	Joukkoistaminen	3
2.1.2	Osallistaminen arkistossa	5
2.1.3	Sosiaalinen metadata kirjastoissa, arkistoissa ja museoissa	6
2.2	Arkistojen avoimuus.....	7
2.3	Joukkoistamisen keskeisiä kysymyksiä.....	10
2.3.1	Arkiston avoimuus ja auktoriteetti	10
2.3.2	Joukkoistamisen työtehtäviä muistiorganisaatioissa	12
2.3.3	Arkiston ja käyttäjän motivaatiot joukkoistamisessa	13
2.4	Yhteenvedo viitekehystä	14
3	TUTKIMUSMENETELMÄ	16
3.1	Tutkimuskysymykset.....	16
3.2	Tutkimusmenetelmä	17
3.3	Tutkimuksen toteutus	17
4	TIEDONHAUN ARTIKKELIT	24
4.1	Tiedonhaun tulokset	24
4.1.1	What's on the Menu?.....	24
4.1.2	Database of the Smokies.....	25
4.1.3	Bentham Papers Transcription Initiative (Transcribe Bentham)....	27
4.1.4	Papers of the War Department 1784-1800	29
4.1.5	Waisda?	31
4.1.6	Your Painting Tagger	33
5	TUTKIMUSTULOKSET	36
6	JOHTOPÄÄTÖKSET	40
7	LÄHTEET	44
	LIITE: TUTKIMUSAINEISTONA KÄYTETYT ARTIKKELIT	48

1 JOHDANTO

Muistiorganisaatioiden eli arkistojen, museoiden ja kirjastojen digitoidun aineiston määrä kasvaa koko ajan. Vaikka aineistoja on internetissä saatavilla, niiden löydettävyys ja käytettävyys ovat riippuvaisia niiden koneluettavuudesta ja metadatatista. Mutta riittääkö vähenevien resurssien muistiorganisaatioissa työvoimaa tekemään digitaalisen aineiston koneluettavuutta sekä luettelointia ja sisällönkuvailua? Voisiko digitaalisen aineiston käytettävyyttä ja saatavuutta parantaa *joukkoistamalla* eli antamalla transkription ja metadatan sisällönkuvailun tekeminen vapaaehtoisten käyttäjien tehtäväksi internetin avulla?

Rose Holleyn (2010) mukaan joukkoistamisen avulla voidaan säästää kustannuksia, aikaa ja henkilökunnan resursseja. Joukkoistamisen avulla voidaan hänen mukaansa myös parantaa aineiston saatavuutta ja sisällönkuvailun laatua sekä sitouttaa käyttäjäyhteisöä. Vaikka Holley kirjoittaa joukkoistamisesta kirjastoissa, hänen näkemyksensä ovat sovellettavissa myös arkistoihin, sillä hänen esimerkkiaineistonsa tehtävinä oli sisällönkuvailua ja transkriptioita.

Joukkoistaminen on uusi ilmiö muistiorganisaatioissa. Tässä opinnäytteessä tutkitaan, kuinka arkistoalan työtehtävien joukkoistamista on organisoitu ja kuinka työn laatua kontrolloidaan sekä sitä, mitkä seikat motivoivat arkistoa ja käyttäjiä joukkoistamiseen. Joukkoistaminen liittyy arkistojen avautumisen jatkumoon, ja opinnäytteessä johdetaan arkistojen avautumiseen liittyvien käsitteiden kautta joukkoistamiseen.

Tutkimuksen menetelmänä on systemaattinen kirjallisuuskatsaus. Tutkimuksella pyritään löytämään tekijöitä, jotka motivoivat organisaatioita ja toisaalta käyttäjiä joukkoistamiseen sekä tekijöitä joukkoistamishankkeiden laadukkuudelle ja laadun kontrollin onnistumiselle, mutta myös epäonnistumisille. Tavoitteena on selvittää hyviä käytäntöjä.

Joukkoistamishankkeita on ja on ollut runsaasti eri puolilla maailmaa, mutta analyttisiä artikkeleita vertaisarvioituissa tieteellisissä julkaisuissa joukkoistamisen organisoinnista ja siihen liittyvästä laadun kontrollista on vähemmän. Tarkastelun kohteeksi valittiin ta-

paustutkimukset, joissa tutkittiin arkistojen, kirjastojen ja museoiden aineiston sisällönkuvailun sekä transkriptioiden joukkoistamista. Julkaisuajankohta rajattiin koskemaan viittä edellistä vuotta eli vuosia 2012-2016.

Toisessa pääluvussa esitellään tutkimuksen viitekehys ja käsitteet. Joukkoistaminen asetetaan arkistojen avautumisen sekä asiakkaiden osallistamisen jatkumoon. Yhteiskunnan muutoksen mukana myös arkiston asiakkaan rooli on muuttunut passiivisesta aktiivisemmaksi. Vapaaehtoisten osallistuminen arkistotyöhön on tuonut mukanaan kysymyksen arkistojen auktoriteetista ja aineistojen luotettavuudesta. Sekä vapaaehtoisilla että arkistoilla on oltava motivaatioita joukkoistamiseen, joten myös motivaatioita selvitetään.

Kolmas pääluku käsittää tutkimusasetelman ja tutkimuskysymyksen sekä systemaattisen kirjallisuuskatsauksen tutkimusmenetelmänä ja tutkimuksen toteutuksen. Neljännessä pääluvussa esitellään tiedonhaun tulos eli joukkoistamisen hankkeet. Viidennessä pääluvussa tuodaan esille tutkimustulokset. Lopuksi esitellään johtopäätökset ja pohdinta.

2 TUTKIELMAN VIITEKEHYS

Muistiorganisaatioiden aineiston joukkoistamisen organisoinnista ei ole tehty varsinaisia kirjallisuuskatsauksia (engl. *literature review*). Joukkoistamisesta yleensä sekä osallistamisesta arkistoissa on tehty kirjallisuuskatsaukset, joita esitellään tässä luvussa. Joukkoistaminen käsitteenä tulee määritellyksi kirjallisuuskatsauksissa.

Tutkielman viitekehyksessä joukkoistaminen liitetään erityisesti arkistojen avautumisen jatkumoon, sillä vapaaehtoistyöllä ja asiakkaiden osallistamisella on pitkät perinteet arkistoissa. Yhteiskunnan muutoksen mukana arkistoammattilaisen ja asiakkaan roolit ovat muuttuneet, ja vapaaehtoisten osallistuminen arkistotyöhön on tuonut mukanaan kysymyksen arkistojen auktoriteetista ja aineistojen luotettavuudesta. Tässä luvussa esitellään, miten arkistojen suhdetta ympäröivään yhteisöön on määritelty eri aikoina ja selvitetään, millainen on arkistojen ja muistiorganisaatioiden suhde avoimuuteen, sosiaaliseen mediaan ja osallistamiseen. Lisäksi selvitetään joukkoistamiseen liittyviä käsitteitä ja käsityksiä. Lopuksi taustoitetaan ja määritetään tutkimusaihe sekä tutkimuskysymykset esitetyssä viitekehyksessä.

2.1 Joukkoistaminen ja arkistot

2.1.1 Joukkoistaminen

Joukkoistamiseen liittyvän terminologian vakiintumattomuudesta johtuen englannin kielen termin *crowdsourcing* rinnalla on käytetty termejä *peer production*, *user-powered systems*, *user-generated systems*, *user-generated content*, *collaborative systems*, *community systems*, *social systems*, *social search*, *social media*, *collective intelligence*, *wikinomics*, *crowd wisdom*, *smart mobs*, *mass collaboration*, *human computation* sekä *community-sourcing* ja *micro-volunteering* (Doan ym. 2011, 86; Ridge 2014, 4).

Crowdsourcing käännetään suomeksi yleensä joukkoistamiseksi. Hyvä esimerkki joukkoistamisen avulla tehdystä työstä on Wikipedia (Leon 2014). Tosin suomenkielinen Wikipedia käyttää termiä joukkouttaminen ja tunnistaa joukkoistamisen vaihtoehtoiseksi suomennokseksi crowdsourcing-termille.

Ensimmäisen kerran termiä crowdsourcing käytti Jeff Howe *Wired*-lehteen vuonna 2006 kirjoittamassaan artikkelissa *The rise of crowdsourcing* (Howe 2006). Joukkoistaminen toimintana on ollut käytössä jo vuosisatoja ennen sanan käyttöönottoa: historian aikana on käytetty kansalaisia tiedon kerääjinä. Esimerkiksi *The Oxford English Dictionary* – sanakirjaa täydennettiin suuren yleisön avustuksella vuodesta 1879 alkaen (Ellis 2014, 5; Ridge 2014).

Joukkoistaminen-termin käytöstä on tehty kaksi kirjallisuuskatsausta, jotka käsittelevät joukkoistamista eri tieteen aloilla. Näissä kirjallisuusselvityksissä ei tehty rajausta informaatiotieteisiin tai muistiorganisaatioihin.

Estelles-Arolas ja Gonzalez-Ladron-de-Guevara (2012) määrittivät joukkoistamista käytetyn usein kuvaamaan monenlaista internetissä tapahtuvaa yhteistyötä ja käyttäjien innovointia. Joukon tehtävänä voi olla rutiininomaiset käytännön tehtävät tai luovat ongelman ratkaisut. Joukkoistaja hyötyy saamalla joukon työpanoksessa sen tiedon ja kokemuksen. Mukaan joukkoistamiseen voidaan kutsua joko avoimesti kaikki kiinnostuneet tai rajataan tiettyyn yhteisöön, jolla on tietyt tiedot ja taidot. Joukkoistaminen voi olla myös yhdistelmä edellisistä, jolloin kutsu on avoin, mutta osallistujia kontrolloidaan. Joukkoistamiseen osallistuvan motivaatio osallistua on saada taloudellista hyötyä, sosiaalista hyväksyntää, itsekunnioitusta tai kehittää omia taitojaan. (Estelles-Arolas & Gonzalez-Ladron-de-Guevara 2012.)

Hossainin ja Kaurasen (2015) mukaan joukkoistaminen on osallistavaa toimintaa yleensä internetin välityksellä. Yksilö, instituutio, yleishyödyllinen järjestö tai yritys pyytää avoimella kutsulla erilaisia yksilöitä tekemään vapaaehtoisesti määritellyn tehtävän. Tehtävän vaikeus ja laajuus voi vaihdella ja joukon osallistuminen voi tarkoittaa heidän työpanostaan, rahaa, tietoa ja / tai kokemusta. Tekijä puolestaan voi saada taloudellista vastinetta, sosiaalista hyväksyntää, itsekunnioitusta tai omien taitojensa kehittämistä. (Hossain & Kauranen 2015.)

Tässä opinnäytteessä käytetään termiä joukkoistaminen kuvaamaan toimintaa, jossa tehtävä annetaan internetin kautta joukolle tehtäväksi rajatulla tai avoimella kutsulla.

2.1.2 Osallistaminen arkistossa

Isto Huvila tutkii vuonna 2015 tekemässään kirjallisuuskatsauksessa, miten osallistamista käsitellään arkiston ja asiakirjahallinnon alan kirjallisuudessa ja mitä käytännön vaikutuksia osallistamisella arkistoissa voi olla. Osallistaminen arkistoissa liitetään yleensä digitaalisiin käytäntöihin, mutta osallistamista käsitteenä ei määritelty. Huvila käyttää Jenkinsin ym. määritelmää osallistamisen kulttuurista sitoutumisena, johon liittyy jakamisen kulttuuri, epämuodollinen mentorointi, usko työpanoksen merkityksellisyyteen ja tunne sosiaalisesta yhteydestä. (Huvila 2015.)

Käyttäjien osallistamisen mahdollistaa sosiaalinen media. Huvila päätyykin toteamaan osallistamisen olevan demokratiaa arkistoissa: osallistamisen avulla arkiston asema aktiivisena toimijana tulee esille omassa yhteisössä sekä mahdollistaa demokratiaa ja avointa hallintoa. Arkistot voivat osallistamalla antaa äänen toiseuden edustajille. (Huvila 2015.)

Huvila erottaa yhdeksän eri osallistamisen diskurssia riippuen siitä, miten artikkelissa käsiteltiin arkistoa, sen käyttäjiä ja heidän osallistamistaan sekä teknologiaa.

1. osallistaminen web 2.0:n kautta sosiaalisena ja teknisenä mahdollisuutena, mutta myös uhkana arkistoja avattaessa yleisölle
2. arkistonhoitajat aktiivisina osallistujina organisaatioissa sekä kokoelmien kokoomisissa
3. arkistonmuodostajat osallistujina
4. ”toiset”, eli suuri yleisö ja amatööritutkijat, informanteina
5. osallistaminen arkistojen uudenaikaisessa, ulkoisessa digitaalisessa käytössä
6. ”toiset” arkistonhoitajana osallistuen arkiston hallintoon ja kuvailuun vertaisina tai konsultteina
7. arkiston asema yhteisössä
8. painopiste toisissa ja ”toisien” osallistaminen
9. ei-osallistaminen

Huvila (2015) määrittää joukkoistamisen liittyvän osallistamisen diskursseihin listan kohtaan 4, ”toisiin” informantteina ja arkistonhoitajina. Joukkoistamisen ja osallistamisen käsitteet nivoutuvat samaan keinoon sitouttaa osallistajat merkitykselliseen yhteyteen kokoelmien kanssa. Tältä osin myös diskurssit 1,2,3,5,6,7 ja 8 voidaan tulkita sekä Estelles-Arolasin ja Gonzalez-Ladron-de-Guevaran että Hossainin ja Kaurasen määritelmien mukaan joukkoistamiseksi. Toisaalta Huvilan määritys joukkoistamiselle lähenee suppeampaa käsitettä *nichesourcing*, jolla tarkoitetaan valikoidulle, asiantuntevalle yleisölle kohdennettua joukkoistamista (esim. De Boer ym. 2012).

2.1.3 Sosiaalinen metadata kirjastoissa, arkistoissa ja museoissa

Karen Smith-Yoshimura johti laajaa tutkimushanketta *sosiaalisen metadatan* käytöstä kirjastoissa, arkistoissa ja museoissa. Sosiaalinen metadata on termi, joka sivuaa myös joukkoistamista. Karen Smith-Yoshimura ja Cyndi Shein (2011) määrittävät sosiaalisen metadatan käyttäjien tekemäksi tägitykseksi, kommentoinniksi, arvioiksi, suosituksiksi tai arvosteluiksi, joiden avulla aineiston muut käyttäjät saavat lisätietoa aineistosta. Tutkimus jakautuu kolmeen osaan: sivustojen kartoitukseen, sivustojen hallinnoijille kohdistettuun kyselytutkimukseen sekä näiden perusteella tehtyyn suosituksiin käytännön toteutukseksi. Tutkimus ei ole kirjallisuuskatsaus vaan se toteutettiin kartoittamalla kirjastojen, arkistojen ja museoiden verkkosivut, joilla käyttäjät osallistuivat muistiorganisaatioiden metadatan kuvailuun (Smith-Yoshimura & Shein 2011, 10.) Tutkimuksessa oli mukana 76 eri organisaatioiden verkkosivustoa, jotka valikoitiin maantieteellisesti edustavasti. Organisaatiot olivat yhteistyössä yhteisönsä kanssa ja niillä oli monipuolisesti käytössä sosiaalinen media. Käyttäjyhteistyön tavoitteena oli mm. kuvailun parantaminen, aiheen määrittely, subjektiiviset arviot, sivuston markkinointi, tutkimuksen helpottaminen ja jakaminen sekä yhteistyön kehittäminen. Tutkijat arvioivat sivustoja käyttäjän näkökulmasta, ei hallinnoijan. Sivustoista kuvailaan mm. kohderyhmä, rekisteröinti, tehtävä sekä se, kontrolloitiinko käyttäjien työtä. Käyttäjyhteistyön tekemän työn laatua ei tutkittu. (Smith-Yoshimura & Shein 2011, 86-166.)

Tutkimusryhmään kuulunut Carol Jean Godby jakaa sivustojen hallinnoinnin kontrollointitavat kolmeen eri ryhmään: ennen julkaisua tarkistettuihin, roskapostin poistamiseen ja

käyttäjien ohjeistamiseen. Väärinkäytöksiä ehkäistiin edellyttämällä esim. käyttäjien rekisteröitymistä, CAPTCHA-fraaseilla sekä ennen julkaisua tapahtuvaan hallinnoijan tekemään tehtävän tarkistamista. (Smith-Yoshimura ym. 2011, 32.)

Helice Koffler ja Karen Smith-Yoshimura selvittävät sivuston hallinnoijan erilaisia tehtäviä ja havaitsivat, että sivuston hallinnointi vei paljon aikaa. Hallinnoijan tehtäviä olivat mm. sivuston tarkkailu, käyttäjien kysymyksiin vastaaminen, sivuston sisällön lisääminen, yhteisöllisyyden kehittäminen ja keskusteluihin osallistuminen. (Smith-Yoshimura ym. 2011, 34.)

Tutkimuksen kolmas osa sisältää yhteenvedon kartoituksesta ja kyselystä luoden niiden pohjalta suositukset sosiaalisen metadatan tekemiseen kirjastoissa, arkistoissa ja museoissa. Smith-Yoshimura ym. arvioivat yhdessä sivustojen hallinnoijien kanssa tekijöitä, jotka vaikuttivat sosiaalisen metadatan onnistumiseen: motivaatio sivun luomiseen, sivun hallinnointi, teknologia ja onnistumisen arviointi. Tutkijat päätyvät suosittelemaan mm. selvien tavoitteiden määrittelyä sosiaalisen median käytössä, käyttäjien motivointia, avoimen lähdekoodin ohjelmistojen käyttöä, käyttäjien tekemän sisällön indeksointia ja hakukelpoisuutta joko erikseen tai yhdessä organisaation sisällön kanssa. (Smith-Yoshimura & Holley 2012, 9-17.)

Smith-Yoshimuran tutkimuksessa ei käytetä joukkoistamista terminä, mutta tutkimusta voidaan soveltaa tähän opinnäytteeseen sikäli, että tutkimuksen kohteena oli käyttäjäyhteistyö internetin kautta, yhtenä muotona muistiorganisaatioiden aineiston tägitys.

2.2 Arkistojen avoimuus

Arkistot ovat paitsi ihmiskunnan muisti, myös vallankäytön väline. Abraham Lincolnin sanoja Gettysburgissa vuonna 1863 mukaellen *“That this nation, under God, shall have a new birth of freedom; and that government of the people, by the people, and for the people, shall not perish from the earth”* Eric Ketelaar soveltaa Lincolnin viestiä arkistoille sopivaksi: *“Archives of the people, by the people, for the people”*. Ketelaar tarkoittaa tällä arkistojen avoimuutta ja saatavuutta, mutta myös ihmisten oikeutta valvoa ja kontrolloida hallintoa. Ihmisten ja kansojen arkistot ovat ihmisiä varten ja ihmisten säilyttämiä, joten niiden olisi oltava avoimia ihmisille. (Ketelaar 1992.)

Arkistojen asema ja tehtävä yhteisössä on muuttunut yhteiskunnan muutosten mukana. Terry Cook määrittelee arkistoille ja niiden suhteelle ympäröivään yhteisöön neljä paradigmaa. Arkistojen tehtävänä oli satojen vuosien ajan toimia todistusaineiston säilytyspaikkana. 1900-luvulle tultaessa arkistot nähtiin myös yhteisön muistina ja arkistojen käyttö uudella tavalla mahdollistui: arkistot säilyttävät muistoja, mutta myös luovat niitä. Muistojen nähtiin olevan valikoituja arkistoissakin, joten seurasi näkemys arkistojen moninaisista totuuksista. Arkistojen tehtäväksi tuli kerätä monenlaisten arkistonmuodostajien aineistoja ja siten edesauttaa identiteetin luomista moniäänisessä yhteiskunnassa. Arkistonhoitajalla oli oltava aktiivinen rooli monipuolisen aineiston hankkimisessa arkistoon, ja varmistaa esimerkiksi vähemmistöjen näkyvyys arkistoissa. Viimeisenä vaiheena on tullut näkemys yhteisön osallisuudesta konkreettiseen arkistotyöhön. (Cook 2013.) Tähän viimeiseen vaiheeseen ajoittuu myös internet, sosiaalinen media ja joukkoistaminen.

Myös Elizabeth Yakel (2011) kirjoittaa arkistojen avautumisesta, mutta hän jakaa muutoksen kahteen vaiheeseen: arkistojen “ensimmäinen suuri avautuminen” oli 1960-luvulla, jolloin pääsy arkistoon avautui tutkijoiden lisäksi suurelle yleisölle. Arkistonhoitajan rooli aineistokeskeisestä asiakirjojen suojelijasta muuttui käyttäjäkeskeiseksi asiakkaan palveluksi. Arkiston asiakkaita ei enää arvotettu, vaan arkistot olivat avoimia kaikille ja kokoelmapolitiikka muuttui myös arkipäivää ja vähemmistöjä huomioivaksi. Jos arkistojen asiakaspalvelu ei olisi muuttunut, vaarana olisi ollut, että asiakkaat olisivat hakenneet informaation toisaalta tai muuttaneet projektejaan aineiston saatavuuden mukaan. “Toinen suuri avautuminen” tapahtui Yakelin mukaan internetin ja Web 2.0 -teknologian avulla. Niiden avulla käyttäjä ei päässyt vain arkistoon vaan suoraan arkistossa oleviin asiakirjoihin. Internetin avulla tutkijat pääsivät tutkimaan suoraan aineistoja eikä arkistonhoitajaa välttämättä tarvittu enää lainkaan. Asiakkaat ja arkistot tai pelkästään asiakkaat keskenään voivat olla vuorovaikutuksessa ja yhteistyössä kommentoimalla, tágittämällä, arvioimalla, luomalla uutta sisältöä ja kuvailua sekä linkittämällä. (Yakel 2011.)

Joukkoistamiseen liittyy keskeisesti käsite *arkisto 2.0*. Li Liew'n ym. mukaan termejä web 2.0, kirjasto 2.0, arkisto 2.0 ja museo 2.0 käytetään kuvaamaan muutosta yksisuuntaisesta internetin staattisen sisällön passiivisesta käytöstä osallistavaan käyttöön, jolloin raja tiedon tuottamisen ja käyttämisen välillä hämärtyy. Keskeisenä on sosiaalisen me-

dian sovellusten kuten blogien, wikien, facebookin ja twitterin käyttö. Arkisto 2.0:n tavoitteeksi Li Liew ym. määrittelevät avoimuuden, yhteistyön ja käyttäjakeskeisen lähtökohdan arkistotyöhön. (Li Liew ym. 2015.)

Kate Theimer tarkentaa, että arkisto 2.0 -käsitteessä ei ole kyse vain arkistojen sosiaalisesta ja teknologisesta muutoksesta, vaan ennen kaikkea asenteellisesta muutoksesta. Arkistojen avoimuus ja toiminnan läpinäkyvyys auttavat tekemään arkistoista käyttäjakeskeisiä. Arkistoammattilaisten rooli on Theimerin mukaan muuttunut portinvartijasta mahdollistajaksi. Arkistotyö on siirtynyt fyysisistä tiloista verkkoon ja käyttäjien tavoitellaan sosiaalisen median avulla, myös uusia ja uudenlaisia käyttäjiä. Internet on myös lisännyt yhteistyötä ja muutosta toimintoihin, esimerkiksi työtehtävistä on tullut toistuvia, niihin palataan ja niitä muokataan. Internetin myötä standardien merkitys on arkistoissa kasvanut. (Theimer 2011, 60-65.)

Cookin ja Yakelin linjoilla on myös Max J. Evans. Vaikka Evans ei käytä termiä joukoistaminen, hän tutkii yleisön osallistamista internetin kautta arkistotyöhön. Hänen mukaansa aineiston olisi oltava avoimesti käytettävissä teknologian avulla, mutta myös löydettävissä erityisesti sisällönkuvailun kautta. Evans näkee ristiriidan arkistojen kasvun ja suuren yleisön odotusten välillä siitä, että kaikki arkistomateriaali olisi löydettävissä luettelointina internetissä. Aineisto voi olla digitoituna, mutta sellaisessa muodossa, että tiedonhaku ei ole mahdollista. Ratkaisuna tähän ei ole vain raha, vaan ennen kaikkea arkistonhoitajien asennemuutos heidän roolistaan ja työstään. Evans kehottaa avaamaan kokoelmat ja aloittamaan yhteistyön sekä muiden organisaatioiden että yleisön kanssa. Evans näkee yleisön mahdollisuutena ja ehdottaa massadigitoimisen jälkeen yleisön mukaan ottamista myös aineiston luettelointiin, jolloin joukkoäly tulisi yhteen arkistoalan toimijoiden kanssa. Tämä yhteistyö vaatisi organisaatioiden toimintatapojen muuttamista. Evans puhuu laajentuvasta metadatatista, johon käyttäjät voisivat lisätä asiasanoja, avainsanoja ja kommentteja arkistonhoitajan koordinoimissa vapaaehtoisten työtä. Evans näkee arkistot aktiivisina toimijoina, jotka voisivat koordinoida kokonaisuutta sekä ohjata ja valvoa käyttäjien luettelointia ja kuvailua. Evansin mukaan avoimuus ja saatavuus ovat kansakunnan arkistojen kulmakiviä. Niitä arkistoja käytetään, joiden kokoelmat ovat avoimia ja saatavilla. Arkistoja käytettäessä kansakunnan sivistys kasvaa ja arkistot saavat ymmärrystä ja arvostusta, kun aineistot tulevat suuren yleisön saataville. Arkistoista voisi tulla tällä tavoin myös osa tiedon luovaa taloutta. (Evans 2007.)

Joukkoistaminen on arkistojen avautumisen jatkumossa toimintaa, jolla suuri yleisö voi konkreettisesti osallistua arkiston työhön, mutta ennen kaikkea tuoda moniäänisyyttä arkistoaineistoon, sekä sisältönä että sisällönkuvailuna.

2.3 Joukkoistamisen keskeisiä kysymyksiä

2.3.1 Arkiston avoimuus ja auktoriteetti

Arkistojen avoimuuteen ja joukkoistamiseen liittyy arkiston auktoriteetin sekä arkistonhoitajan ja asiakkaan roolien muutos. Arkistonhoitajan ja arkistonkäyttäjän roolien muuttuminen vaikuttaa moneen asiaan. Arkiston työtehtäviä on tehnyt arkistonhoitaja, jolla on auktoriteetti. Gummeruksen uuden suomen kielen sanakirja määrittelee auktoriteetin ”arvovaltaiseksi ja asiantuntevaksi henkilöksi”. Auktoriteetin käsite liittyy olennaisesti arkiston avoimuuteen ja joukkoistamiseen siinä vaiheessa, kun käyttäjät tekevät arkistoalan töitä joukkoistamalla.

Elizabeth Yakelin (2011) mukaan Web 2.0 -teknologia on mahdollistanut sosiaalista *vertaistuotantoa*. Yakel käyttää englanninkielisen Wikipedian määritystä vertaistuotannosta (engl. *peer production*) ”itseohjautuvan yhteisön tavaksi tuottaa palveluita tai hyödykkeitä koordinoitusti”. Yakel täsmentää vertaistuotannon sisältävän vastuun ja toiminnan yhteisön puolesta sekä ajatuksen yhteisestä tavoitteesta. Asiakirjojen autenttisuutta perustellaan usein arkistojen institutionaalisella auktoriteetillä. Institutionaalisen auktoriteetin lisäksi arkistoilla on myös kognitiivista auktoriteettia, joka on kokemuksesta saatua tietoa pätevydestä ja siten saadusta luottamuksesta. Yakelin mukaan arkistojen maine, luottamus ja identiteetti toimivat kognitiivisena auktoriteettinä. Vertaistuotanto ei pohjautu hierarkiaan eikä auktoriteettiin, vaan kognitiivisen auktoriteettiin. Yakel nostaa esille kysymyksiä auktoriteetista, jos metadataa avataan vertaistuotannolle: Kuinka paljon auktoriteetin suoja voi antaa vertaistuotannolle arkiston auktoriteetin nimissä? Voivatko arkistot säilyä luotettuina instituutioina, jos ne jakavat omaa auktoriteettiaan? Voiko arkistoaineiston metadata olla moniääninen tai sisältää ristiriitaista tietoa? (Yakel 2011.)

Kognitiivisen auktoriteetin ylläpitämisen keinoiksi Yakel näkee erottelun ja tarkastamisen ennen julkaisua. Erottelussa käyttäjille on oma erillinen tila toimia, esimerkiksi käyttäjien lisäämä tieto eriytetään arkiston omasta sisällöntuotannosta. Käyttäjien tuottaman sisällön oikeellisuutta ja luotettavuutta voidaan valvoa hallinnoimalla. Voidaan käyttää myös vastuuvapauslauseketta “sivuston ylläpitäjä ei vastaa sivuston oikeellisuudesta”, jolloin käyttäjien sisältö erotellaan virallisesta sisällöstä. Käyttäjä voi tällöin päätellä, mikä tieto on auktorisoitua. Yakelin mukaan arkistonhoitaja ei ole vertaistuotannossa yhteisön jäsen vaan erillinen, kontrolloiva toimija. Arkistonhoitajan täytyy jakaa omaa auktoriteettiään yhteisölle ja jakaa vastuuta sivuston hallinnoinnille. (Yakel 2011.)

Myös Huvilan kirjallisuuskatsauksen mukaan yksi osallistamisen tavoista on arkistonhoitajien rooli tehdä yhdessä käyttäjien kanssa, ei käyttäjiä varten. Käyttäjiksi ja osallistujiksi Huvilan tutkimus määrittelee asiakirjojen luojat, paikallishistorian harrastajat ja muut informantit, jotka voivat osallistua esimerkiksi sisällönkuvailijoina, valokuvien tunnustajina, transkriptioiden tekijöinä. Huvila nostaa esille kysymyksen arkistotyön uskottavuudesta: pitäisikö ammattilaisen työ erotella osallistujan työstä? (Huvila 2015.)

Bonney ja Simon menevät Oomenin ja Aroyon mukaan pitemmälle auktoriteetin jaossa eritellessään ammattilaisten ja amatöörien rooleja kulttuuriperinnön joukkoistamisessa (Bonney & Simon. Tässä Oomen & Aroyo 2011).:

- arkistonhoitajat suunnittelevat ja käyttäjät toteuttavat.
- arkistonhoitajat suunnittelevat ja käyttäjät analysoivat, hiovat suunnitelmaa tai jakavat aineistojaan.
- arkistonhoitajat ja käyttäjät suunnittelevat ja tekevät yhteistyötä jokaisessa vaiheessa.
- arkistonhoitajat vain isännöivät hanketta ja käyttäjät voivat työstää vapaasti aineistoa.

Haasteena on siis tehdyn työn laadun taso. Aiemmin instituutiot vastasivat itse työnsä laadusta. Joukkoistamisessa työn tekijät eivät välttämättä ole ammattilaisia tai asiantuntijoita vaan todennäköisemmin amatöörejä. Oomen ja Aroyo antavat kolme keinoa laadun kontrollointiin: toimintaohjeet, halutun laatutason “imago” sekä virheellisen tiedon suodattaminen tai korjaus. (Oomen & Aroyo 2011.)

Keskeisiä teemoja tässä opinnäytteessä ovat auktoriteetti, joukkoistetun työn laatu ja laadun kontrolli. Laatu voidaan kontrolloida erottelemalla joukkoistamalla tehty työ ammattilaisen työstä ja/tai ohjeistamalla ja tarkastamalla joukkoistamalla tehty työ ennen julkaisua. Ammattilaisen rooli joukkoistamisessa ilmenee siinä, missä laajuudessa hän tekee yhteistyötä käyttäjien kanssa ja jakaa omaa auktoriteettiä.

2.3.2 Joukkoistamisen työtehtäviä muistiorganisaatioissa

Mia Ridgen mukaan muistiorganisaatioiden kulttuuriperinnön joukkoistamisprojekteissa yleisö tekee tehtäviä, joita ei voi tehdä automaattisesti koneellisesti (Ridge 2014, 2). Sekä Ridge että Oomen ja Aroyo määrittelevät kulttuuriperinnön joukkoistamisen avulla tehdyiksi töiksi formaatin siirron toiseksi kuten transkriptiot, kohteen kuvailun tágien ja luokituksen avulla ja uuden tiedon lisäämisen tai luomisen (Oomen & Aroyo, 2011; Ridge 2014, 6). Lisäksi Oomenin ja Aroyon mukaan muistiorganisaatiot voivat yhteistyössä käyttäjien kanssa tehdä yhteiskuratointia esimerkiksi toteuttamalla näyttelyn yhteistyössä harrastaja-asiantuntijan kanssa tai hakea rahoitusta joukkorahoituksen avulla (Oomen & Aroyo, 2011).

Tieteen termipankki määrittelee transkription ”tekstistä toiseen dokumenttiin tehdyksi jäljennökseksi”. Esimerkiksi käsikirjoituksena säilyneen tekstin tai historiallisen dokumentin sisällön kirjoittaminen tiedostoksi tuottaa transkription.

Useisiin tutkimuksiin viitaten Petek määrittää tágityksen toiminnaksi, jonka avulla käyttäjä organisoi informaatiota itselleen merkityksellisellä tavalla. Tágit ovat vapaasti valittuja avainsanoja, jotka kuvaavat kohdetta ja auttavat löytämään sen myöhemmin. Tági on keskeinen Web 2.0:n työväline. Tágien kokonaisuutta kuvailemaan käytetään termiä *folksonomia*, joka on sosiaalista luokitusta ilman hierarkiaa ja rakentuu vapaasti kasvavaksi sanojen verkostoksi. Petek tutki Flickr-kuvasivuston käyttäjien motivaatiota tágitykseen ja vertasi tágitystä ammattilaisten tekemään asiasanoitukseen. Käyttäjien motivaatiot olivat useimmiten henkilökohtaisia, harvemmin perusteena oli yhteinen hyvä. Käyttäjät antoivat tägejä enemmän kuin ammattilaiset asiasanoja. (Petek 2012.)

Gorzalski viittaa useisiin tutkimuksiin, joiden mukaan tägitys parantaa aineiston löydettävyyttä. Ongelmaksi voi tulla käyttäjien metadataan tekemän sisällön, tägituksen harhaanjohtavuus tai ristiriitaisuus suhteessa sisältöön tai muihin tägeihin. (Gorzalski 2013.)

Yleisimpiä joukkoistamisen tehtäviä muistiorganisaatioissa ovat transkriptiot ja tägitys, jotka siten määrittyivät myös tutkielman kohdeaineistoksi.

2.3.3 Arkiston ja käyttäjän motivaatiot joukkoistamisessa

Aiemmin ammattilaisilla oli täysi kontrolli informaation laatuun ja oikeudellisuuteen hankinnassa, järjestämisessä ja kirjaamisessa, mutta joukkoistamisen avulla voidaan rakentaa kulttuuriperintöä yhteistyössä. Joukkoistamisessa on tärkeää motivointi ja lojailien käyttäjien saanti. Oomenin ja Aroyon mukaan sosiaalista motivointia voi olla yhteydenpidon ja jäsenyyden kautta sekä jakamalla. (Oomen & Aroyo 2011.)

Oomenin, Gligorovin ja Hildebrandin (2014) mukaan motivaationa käyttäjälle osallistua joukkoistamiseen on kiinnostus alaan tai aiheeseen. Amatöörejä motivoi innostus aiheeseen, ei raha. Myös Huvilan osallistamiseen liittyvässä tutkimuksessa ilmenee se, että käyttäjiä motivoi oma kiinnostus aiheeseen ja erilaiset kannustimet, ei niinkään arkiston tehtävät. Monet arkistonhoitajien osallistamisen aloitteista liittyvät osallistujien alakulttuurin harrastajien kaltaiseen faniuteen. (Huvila 2015.)

Rose Holleyn (2010) mukaan käyttäjiä motivoi joukkoistamisessa mm. mielenkiinto aiheeseen, velvollisuuden tunto yhteisölle ja halu olla mukana tavoitteen saavuttamisessa. Motivaatiota osallistua joukkoistamiseen voitiin saada lisäämällä tehtäviä säännöllisesti, kasvattamalla haastetta, käyttämällä virtuaaliyhteisön foorumeita, ohjeistamalla selkeästi, palkitsemalla esim. julkisella ranking-listalla tai todistuksilla. Myös mittari, jolla ilmaistiin tavoitteen saavuttamista lisäsi motivaatiota. (Holley 2010.)

Arkistoille ja muille muistiorganisaatioille on omat motivaationsa joukkoistamiseen. Johan Oomen, Rite Gligorov ja Michiel Hildebrand erottavat neljä eri motivaatiota (Oomen, Gligorov & Hildebrand 2014):

1. Joukkoistaminen kaventaa “semanttista kuilua” ammattilaisten käyttämän terminologian ja käyttäjien hakutermien välillä. Ammatillaiset käyttävät sanastoja ja tesaurusuksia, kun taas yleisölle ne ovat vieraita eivätkä ole siten hyödynnettävissä.
2. Joukkoistamisen avulla saadaan lisätietoa ja kontekstia. Käyttäjien merkinnät voivat muodostaa tärkeän osan muistiorganisaatioiden tiedonhallinnalle ja luettelointiprosesseille. Digitaalisilta aineistoilta puuttuu usein metadata, joka vaikeuttaa niiden löydettävyyttä. Organisaatioilla ei ole resursseja lisätä manuaalisesti metadataa. Sen tähden muistiorganisaatiot etsivät vaihtoehtoisia tapoja tuottaa täydentävää metadataa, käyttämällä esim. puheentunnistusta ja yhteistyötä käyttäjien kanssa.
3. Joukkoistamisen avulla saadaan resursseja muistiorganisaatioiden tulevaan työvirtaan. Esim. käyttäjät täydentävät syntyjään digitaalisten aineistojen puuttuvaa metadataa.
4. Joukkoistamisen avulla lisääntyy organisaation ja yleisön yhteys.

Aiempien tutkimusten perusteella käyttäjien motivaationa joukkoistamiseen on oma kiinnostus aiheeseen sekä halu osallistua yhteiseen asiaan. Käyttäjiä voidaan motivoida monin tavoin: aktiivisella yhteydenpidolla, selkeällä ohjeistuksella, tavoitteilla ja palkitsevilla. Organisaation motivaationa voi olla lisäresurssien, käyttäjän näkökulman ja lisätiedon saaminen sekä yhteyden saaminen käyttäjiin.

2.4 Yhteenveto viitekehystä

Tässä opinnäytteessä käytetään termiä joukkoistaminen kuvaamaan toimintaa, jossa tehtävä annetaan internetin kautta joukolle tehtäväksi rajatulla tai avoimella kutsulla. Tehtävän laajuus ja vaikeus voi vaihdella.

Tutkimuksen kohteena oleviksi työtehtäviksi määriteltiin yleisimmät joukkoistamisen työtehtävät, transkriptiot ja sisällönkuvailu.

Joukkoistamalla organisaatio voi saada lisätietoja aineistosta käyttäjien tiedon ja kokemuksen kautta. Muistiorganisaatioiden aineiston näkyvyyttä ja siten löydettävyyttä voidaan parantaa joukkoistamalla. Joukkoistamisen motivaationa organisaatiolle voi olla myös ”semanttisen kuilun” kaventaminen, lisätiedon saaminen, resurssipulan paikkaaminen ja organisaation yhteys yleisöön.

Käyttäjien motivaationa osallistua joukkoistamiseen voi olla mielenkiinto aiheeseen, velvollisuuden tunto yhteisölle tai halu olla mukana tavoitteen saavuttamisessa. Joukkoistamisessa voidaan motivoida käyttäjiä lisäämällä tehtäviä säännöllisesti, haastetta kasvatamalla, selkeällä ohjeistuksella, palkitsemalla ja havainnollistamalla tavoitteen saavuttamista.

Luotettavuuden takaamiseksi joukkoistamalla tehdyn työn laatua on kontrolloitava. Joukkoistamisen avulla tehdyn työn laadun kontrolloinnin keinoja ovat ohjeistus, suodatus tai virheellisen tiedon korjaus.

Keskeisiä teemoja tässä opinnäytteessä ovat auktoriteetti, joukkoistetun työn laatu ja laadun kontrolli. Laatua voidaan kontrolloida erottelemalla joukkoistamalla tehty työ ammattilaisen työstä, ohjeistamalla, tarkastamalla joukkoistamalla tehty työ ennen julkaisua. Ammattilaisen rooli joukkoistamisessa ilmenee siinä, missä laajuudessa hän tekee yhteistyössä käyttäjien kanssa.

3 TUTKIMUSMENETELMÄ

Tässä luvussa esitellään systemaattinen kirjallisuuskatsaus metodina ja luodaan sen pohjalta tutkimussuunnitelma. Tutkimuksen prosessi kuvataan sekä määritellään rajaukset ja teemat tutkimusaiheelle.

3.1 Tutkimuskysymykset

Tutkielmassa etsittiin vastausta seuraavaan kysymykseen:

Miten muistiorganisaatioiden eli museoiden, arkistojen ja kirjastojen aineiston sisällönkuvailua ja transkriptiota voidaan joukkoistaa eli organisoida käyttäjien tehtäväksi?

Muistiorganisaatioiden aineistoon liittyvät työtehtävät, joita voidaan joukkoistaa, rajattiin tutkielmassa koskemaan transkriptiota ja sisällönkuvailua. Näitä työtehtäviä voivat tehdä sekä asiantuntijat että suuren yleisön edustajat, ja sisällönkuvailuun organisaatio saa kahden eri käyttäjäkunnan näkökulmaa.

Joukkoistamisen avulla tehdyn sisällönkuvailun ja transkription organisointiin liittyy kysymyksiä auktoriteetista, luottamuksesta, työn laadusta sekä motivaatiosta. Vastausta näihin etsittiin selvittämällä rekrytoinnin kohdentamista ja tarkastelemalla annettuja työtehtäviä sekä työn laadun kontrollia. Lisäksi tutkittiin ammattilaisen eli joukkoistamishankkeen hallinnoijan ja tehtävän tekijän eli käyttäjän rooleja sekä organisaation ja käyttäjän motivaatioita joukkoistamiseen.

Tutkielman tavoitteena oli löytää hyviä käytänteitä joukkoistamisen avulla tehtyjen arkistoalan töiden organisointiin. Tekniset ratkaisut rajataan pois opinnäytteen aihepiiristä.

3.2 Tutkimusmenetelmä

Systematic literature review suomennetaan *systemaattiseksi kirjallisuuskatsaukseksi*. Toisin termin *review* käänös *katsaukseksi* on tässä yhteydessä harhaanjohtava, koska kyseessä ei ole yhteenveto vaan kriittinen analyysi. (Salminen 2011, 5.)

Kirjallisuuskatsauksen avulla voi muodostaa kokonaiskuvan tietystä aihealueesta, mutta sen avulla voi myös tunnistaa ristiriitaisuuksia ja ongelmia. Kirjallisuuskatsauksen voi tehdä tietyn tieteenalan näkökulmasta tai se voi olla poikkitieteellinen. Olennaista on se, mihin tarkoitukseen kirjallisuuskatsaus tehdään. (Suhonen, Axalin & Stolt 2016, 7.) Objektiiivista tutkimuskatsausta tehdessä tutkitaan systemaattisesti kaikki lähteet sekä kuvataan ja perustellaan se, mitä ja miten on tehty. Tällä tavalla toimimalla mahdollistetaan tutkimuksen toisintaminen, systemaattisuus, eksplisiittisuus ja kattavuus. (Fink 2014, 14-15.) Systemaattisessa kirjallisuuskatsauksessa tutkitaan myös käytettyjen lähteiden keskinäistä suhdetta sekä menetelmää, jolla siteeratut tulokset on hankittu. Systemaattisen kirjallisuuskatsauksen voi sanoa olevan tutkimusta tutkimuksesta. Systemaattisessa kirjallisuuskatsauksessa tietyn asiakokonaisuuden laaja, jo olemassa oleva tutkimusaineisto käydään läpi poimien poikkeavuudet ja yhteneväisyydet tiiviiseen kokonaiskuvaan. Tämä asetetaan sekä historialliseen että oman tieteenalansa kontekstiin ja siten testataan hypoteeseja ja tutkimustulosten johdonmukaisuutta. Kirjallisuuskatsauksen avulla voidaan kehittää tai arvioida olemassa olevaa teoriaa, mutta myös kehittää uutta teoriaa. Kirjallisuuskatsauksen avulla voidaan tunnistaa ongelmia tai kuvata tietyn teorian historiallista kehitystä. (Salminen 2001, 3-4, 9.)

3.3 Tutkimuksen toteutus

Arlene Fink (2014, 4-5) on jakanut systemaattisen kirjallisuuskatsauksen laatimisen seitsemään vaiheeseen:

1. Tutkimuskysymyksen asettaminen
2. Tietokantojen valinta
3. Hakulauseiden muotoilu
4. Rajausten määrittely 1

5. Rajausten määrittely 2
6. Katsauksen teko
7. Tulosten analyysin ja synteesin teko

Seuraavaksi tarkastellaan näitä vaiheita yksitellen ja esitellään, kuinka niitä sovellettiin tutkielman systemaattista kirjallisuuskatsausta tehtäessä.

1. Tutkimuskysymyksen asettaminen

Ensimmäiseksi määriteltiin tutkimuskysymys, jolla rajataan tutkimuksen aihe. Tutkimuskysymykseksi määriteltiin:

Miten muistiorganisaatioiden eli museoiden, arkistojen ja kirjastojen aineiston sisällönkuvailua ja transkriptiota voidaan joukkoistaa eli organisoida käyttäjien tehtäväksi?

Joukkoistamisen tutkimus rajattiin tarkemmin koskemaan annetun työn laadun kontrollia:

Miten taataan luottamus joukkoistamisen avulla tehtyjen transkriptioiden ja sisällönkuvailun laatuun?

2. Tietokantojen valinta

Seuraavaksi tunnistettiin aineisto valitsemalla aiheenmukainen kirjallisuus ja tietokannat sekä läpikäymällä lähdeviitteet ja asiantuntijalähteet kirjallisuuden löytämiseksi (Fink 2014, 16). Tiedonhaun tulos vaikuttaa siihen, pitääkö tutkimuskysymystä tarkentaa tai muuttaa.

Tietokannoiksi valittiin Tampereen yliopiston kirjaston käytössä olevat kotimaiset ja kansainväliset tietokannat kesällä 2016, jotka olivat.

Kotimaiset tietokannat:

- Tamcat
- Arto

- Melinda

Kansainväliset tietokannat:

- EBSCOhost Academic Search Premier
- Emerald
- Taylor & Francis Online Journal Library
- Wiley Online Library

Lisäksi mukana oli informaatiotutkimuksen keskeiset viitetietokannat LISA ja LISTA, joiden viitteiden artikkeleita löytyi Nelliportaalin kautta.

3. Hakulauseiden muotoilu

Hakulauseen lisäksi määritettiin, mihin hakusana kohdennettiin: avainsanoihin, sisällönkuvailuun, abstraktiin, otsikkoon, kokotekstiin. Haku kannatti kohdentaa otsikkoon tai abstraktiin, jos hakusana oli yleinen. Tämä kohdentaminen tehtiin eri tietokannoissa eri tavoin. (Lehtiö & Johansson 2016.)

Haun kohdentaminen vaihteli tietokannasta riippuen, sillä hakujen kohdentaminen ei ollut mahdollista kaikissa tietokannoissa. Hakua ei voinut kaikissa tiedonhauissa kohdentaa esim. abstraktiin, jolloin haku kohdennettiin kokotekstiin, jos mahdollista. Kun hakulauseetta tarkennettiin joukkoistamisen avulla tehdyn työn laatuun sekä metadataan, haku täytyi kohdentaa kokotekstiin, koska työn laatu ja metadata eivät olleet artikkeleiden pääaiheina, vaan näitä teemoja käsiteltiin laajemmissa asiayhteyksissä. Haun kohdentamisessa abstraktiin oli havaittavissa harhaanjohtavaa epätarkkuutta, sillä abstraktiin kohdistettaessa haku huomioi myös tekijän tittelin (esim. *metadata librarian*) ja copyright-informaatiotekstin.

Tarvittaessa käytettiin hakutermien määrittelyssä myös asiantuntijoita apuna. (Fink 2014, 20-25.) Finkin mukaan (2014, 27) ongelmana oli usein uusien aiheiden käsitteet ja kuvailu, koska niitä ei käytetä yhdenmukaisesti. Hakulauseke muokattiin kuhunkin tietokantaan sopivaksi esim. sanakatkaisun ja fraasien käytössä (Lehtiö & Johansson 2016, 42-43).

Tietokantojen sekä määrällisesti että laadullisesti vaihtelevaa asiasanojen käyttöä pyrittiin hallitsemaan tiedonhaun systemaattisuudella. Osa artikkeleista oli asiasanoitettu vain kahdella hyvin yleisellä asiasanalla, myös abstraktien laajuus vaihteli. Hakutermien valinnassa ei pyritty hakutulosten kattavuuteen vaan ennemminkin edustavaan valikoimaan alan tutkimuksia. Alan terminologian vakiintumattomuus vaikutti myös artikkeleissa käytettyihin termeihin ja hankaloitti siten kokotekstihakuja. Tämä näkyi opinnäytteen tiedonhaussa, koska muistiorganisaatioiden joukkoistamishankkeisiin liittyvä terminologia ei ollut vakiintunutta suomeksi eikä englanniksi. Siltä osin kuin vaihtoehdoisen termin on katsottu kuvastavan toimintoja, jotka liittyvät joukkoistamisen kaltaiseen toimintaan, se sisällytettiin opinnäytteen tiedonhakuun. Sen tähden käytettiin useita hakusanoja. Suomalaisiin ja kansainvälisiin tietokantoihin määriteltiin erilaiset hakulauseet, koska termien käyttö Suomessa poikkeaa englanninkielisestä eikä ole siten käännettävissä suoraan englannin kielestä suomen kielelle.

Lauseen lisäksi määritettiin, mihin hakusana kohdennettiin: avainsanoihin, abstraktiin vai kokotekstiin. Hakusanan kohdentaminen ei ollut kaikissa tietokannoissa mahdollista. Auktoriteetin jakoa ei näy hakulauseessa vaan sitä voidaan tutkia työtehtävien ja työn laadun kontrollin kautta. Työtehtävät ja työn laadun kontrolli määrittelevät hallinnoijan ja käyttäjän rooleja, jolloin päästään tutkimaan auktoriteettia ja sen jakoa: kuinka hallinnoija luovuttaa omia työtehtäviään käyttäjän tehtäväksi. Myöskään motivointi ei sisälly hakulauseeseen vaan se ilmenee joukkoistamisen lähtökohtana: miksi organisaatio aloittaa joukkoistamisen ja kuinka joukko rekrytoidaan.

Englanninkielinen hakulause:

(crowdsourcing OR participate OR "web 2.0" OR "user generated content" OR cooperation OR collaborative) SU

AND (metadata OR transcription) AB (hakusana "tags" sisältyy "metadataan")

AND (quality OR moderate) TX

Suomenkielinen hakulause:

(joukkoistami? OR osallistami? OR "web 2.0") ASIASANAHAKU

AND (metadata? OR transkriptio? OR tåg?) SANAHAKU

Suomalaisissa tietokannoissa ei ollut käytettävissä kokotekstiä tai abstraktia kuten kansainvälisissä tietokannoissa, joten sanahaku kohdentui suomalaisissa tietokannoissa siten otsikoihin ja asiasanoihin.

4.-5. *Rajausten määrittely*

Jotta tuloksena ei olisi ollut satoja artikkeleita vaan relevantti ja tarkka katsaus, niin aineistoon tehtiin erilaisia rajauksia. Rajauksena eli seulana käytettiin julkaisun kieltä ja julkaisuaikaa. (Fink 2014, 51-52.) Kun hakutulos havaittiin riittämättömäksi, lisää artikkeleita löydettiin keskustelemalla asiantuntijoiden kanssa, tutkimalla luotettavia nettisivuja ja korkealaatuisten tutkimusten lähdeviitteitä. (Fink 2014, 28). Tietokantahakujen lisäksi tehtiin manuaalisia hakuja valittujen artikkelien lähdeluetteloista. Mukaanotto- ja poissulkukriteereillä varmistettiin katsauksen pysyminen tutkimuksen fokuksessa (Niela-Vilén & Hamari, 2016, 26-27). Etukäteen tehtävällä mukaanotto- ja poissulkukriteerien sekä menetelmien tarkalla määrittelyllä vähennettiin katsauksen vinoutumaa (Valkeapää 2016, 56).

1. Rajausten määrittely:

Seuraavaksi seulottiin asettamalla mukaanotto- ja poissulkukriteerit:

- Tutkimus käsittelee muistiorganisaation joukkoistamishanketta analyytisesti.
- Ajanjakso rajattiin vuosiin 2012-2016.
- Tutkimuksen kieli on suomi tai englanti.
 - Tutkimukset ovat artikkeleita monografioissa tai vertaisarvioituissa lehdissä, jotka ovat käytettävissä Tampereen yliopiston kokoelmissa tai tietokannoissa maksuttomina kokoteksteinä (maksuttomuus on poikkeuksellisesti tässä rajauksena, koska kyseessä on opinnäyte).
- Lisäaineistoja etsittiin tutkimalla lähdeviitteitä.

Systemaattinen tiedonhaku rajattiin koskemaan ajallisesti viisi kalenterivuotta taannehtivasti vuosiin 2012-2016 ja kieliksi rajattiin suomi ja englanti.

2. Rajausten määrittely:

Mukaan otetun aineiston sopivuus otettiin tarkempaan sisällön tarkasteluun. Toisena rajauskriteerinä käytettiin aineiston metodologiaa ja arvioitiin tieteellistä laatua (Fink 2014, 77). Selvitettiin, onko tutkimus sekä sisällöllisesti että muodollisesti pätevä ja olivatko tutkimuksen tiedonlähteet luotettavia ja päteviä (Fink 2014, 104-105). Lisäksi tutkittiin, olivatko analyyttiset metodit sopivia tutkimusdatan laadulle ja ominaisuuksille. Tulosten piti olla merkityksellisiä käytännöllisillä ja tilastollisilla termeillä. Tulosten piti olla esitetty järkevällä tavalla, kuvailen tutkimuksen vahvuuksia ja heikkouksia. (Fink 2014, 55, 77.)

Rajausten määrittelyn jälkeen saadut 198 viitettä käytiin läpi, ja niistä valikoitui tarkempaan sisällön tarkasteluun 24 artikkelia. Suomenkielisiä artikkeleita em. rajausten määrittelyllä ei löytynyt. Tätä hakutulosta selittänee aiheen tarkka rajaus ilman hakumahdollisuutta kokotekstiin sekä se, että suomalaisissa tietokannoissa ei ole suomenkielisiä julkaisuja haetusta aiheesta siinä määrin kuin kansainvälisissä. Tietokannoista löytyi osin samoja artikkeleita. Tarkastelun kohteeksi valittiin tapaustutkimukset, joissa tutkittiin muistiorganisaation aineiston tai sen metadatan avaamista joukkoistamiselle.

Artikkelit luettiin kokonaan, jotta hakutuloksista löydettiin haluttua näkökulmaa eli käyttäjien motivaatiota ja työn laatua käsittelevät artikkelit. Hakutuloksista valikoitui kuusi eri joukkoistamishanketta seitsemässä artikkelissa kuvattuna, yhtä hanketta käsiteltiin kahdessa eri artikkelista (kts. LIITE Tutkimusaineistona käytetyt artikkelit). Keskeiseksi teokseksi nousi Mia Ridgen toimittama julkaisu ”Crowdsourcing our cultural heritage”, jossa oli peräti viisi tutkimusaineiston artikkelia.

6. Katsauksen teko

Joukkoistamishankkeista rajattiin pois kontrolloidut koetilanteet sekä yhteen organisaatioon rajatut joukkoistamiset. Tutkimuskohteeksi valikoidun joukkoistamishankkeen täytyi olla avoin laajemmalle joukolle. Hakutuloksen kuusi joukkoistamishanketta esiteltynä seitsemässä eri artikkelissa luettiin systemaattisesti läpi ja analysoitiin. Artikkelit esitellyistä hankkeista eivät olleet yhteismitallisia. Niissä ei käsitelty samoja asioita yhteneväi-

sesti siten, että olisi saatu täysin vertailukelpoiset taulukot. Artikkeleista pyrittiin löytämään ne hankkeet, jotka käsitelivät muistiorganisaation aineiston transkription tai sisälönkuvailun joukkoistamista.

7. Tulosten analyysin ja synteessin teko

Analyysissä järjestettiin ja luokiteltiin aineistoa. Sen jälkeen aineistosta etsittiin yhtäläisyyksiä ja eroja, ryhmiteltiin ja vertailtiin. Samankaltaiset teemat yhdisteltiin, niitä vertailtiin ja nimettiin sisältöä kuvaavasti. Tuloksia tulkittiin siten, että niistä muodostui synteesi, “ymmärrystä lisäävä kokonaisuus”. Analyysi ja synteesi olivat yhtäaikaisia. (Whittemore 2005, Whittemore & Knafl 2005, Aveyard 2007, tässä Niela-Vilén & Hamari 2016, 30-31)

Hakutulosten artikkelit valikoitiin mukaan, jos niissä oli käsitelty seuraavia teemoja:

- Työtehtävä: Mitä hankkeessa tehdään?
- Rekryointi: Onko hankkeissa ollut avoin vai rajattu kutsu joukkoistamiseen?
- Motivaatio: Mikä joukkoa motivoi? Miten joukkoa motivoidaan? Mikä motivoi organisaatiota joukkoistamiseen?
- Työn laatu: Miten laatua kontrolloidaan?
- Auktoriteetin jako: Miten ammattilaisen ja käyttäjän roolit jakautuvat?

Näiden teemojen pohjalta kootaan lukuun 5 taulukot siitä, miten systemaattisen kirjallisuuskatsauksen hakutuloksen artikkeleissa näitä teemoja on käsitelty. Löydetyistä artikkeleista tutkittiin käyttäjien rekryointia, käyttäjien ja muistiorganisaation motivaatiota, käyttäjän ja hallinnoijan rooleja sekä tehdyn työn laatua ja laadun kontrollia. Haettiin yhteisiä piirteitä, mutta myös yksittäisiä havaintoja, joita voitaisiin määritellä hyviksi käytänteiksi. Taulukoiden sisältämät numerot viittaavat luvussa 4 esiteltyihin joukkoistamishankkeisiin. Johtopäätökset ja pohdinta tehdään luvussa 6.

4 TIEDONHAUN ARTIKKELIT

4.1 Tiedonhaun tulokset

Tässä luvussa esitellään edellisessä luvussa kuvaillulla tutkimusmenetelmällä löydetty kuusi joukkoistamishanketta. Kaikki esitellyt hankkeet pitivät joukkoistamista kannattavana ja kirjasivat havaintoja myös tulevia joukkoistamishankkeita varten. Hankkeet olivat erilaisia aineistoltaan ja tehtäviltään, mutta joukkoistamiseen liittyvään organisointiin löytyi haluttuja teemoja. Tutkimuskysymysten perusteella koottiin taulukot, jotka esitellään hankkeiden esittelyn jälkeen luvussa 5.

Hankkeiden esittelyssä termejä on yhtenäistetty käyttämällä arkiston tai muistiorganisaation ammattilaisesta termiä *hallinnoija* ja joukkoistetun tehtävän tekijästä termiä *käyttäjä*.

4.1.1 What's on the Menu?

Ben Vershbow (2013) kirjoitti ensin What's on the Menu -hankkeesta yksin ja sitten yhdessä Michael Lascaridesin (2014) kanssa. New York Public Libraryn 9000 ruokalistan joukkoistamishankkeen ”What's on the Menu?” tehtävänä oli transkriptio eli kirjoittaa digitoituja, kokonaisia ruokalistoja uudelle alustalle. Lisäksi käyttäjä kirjasi metadataan neljä kenttää: ruokalajin, hinnan, valuutan ja määritelmän alkupala-pääruoka-jälkiruoka.

Joukkoistamiskutsu oli avoin. New York Times uutisoi New Yorkin ruokalistojen transkriptioiden joukkoistamishankkeesta, jonka jälkeen käyttäjiä ilmaantui enemmän mukaan. Käyttäjiin pidettiin yhteyttä sähköpostitiedotteen ja sosiaalisen median kautta. Käyttäjiltä ei vaadittu rekisteröintiä tai kirjautumista, koska osallistumisen kynnys haluttiin pitää matalana. Vaikka rekisteröinti olisi mahdollistanut palkitsemisen ja kutsuvierastilaisuudet, niin se nähtiin myös esteeksi.

Käyttäjien työn jälki oli korkealaatuista eikä väärinkäytöksiä havaittu. Käyttäjän tekemä työ meni hallinnoijalle tarkastusvaiheeseen, jossa se oli myös muiden käyttäjien tarkasteltavana ja muokattavana. Sen jälkeen työ siirrettiin tilaan, jossa muokkaus ei ollut enää

mahdollista. Transkription ajan valittu ruokalista oli lukittu tehtävän aloittaneen käyttäjän käyttöön, mutta palautteen perusteella mahdollistettiin usean käyttäjän yhtäaikainen työskentely.

Joukkoistamishanke vaati hallinnoinnin resursseja New York Public Librarylta: ruokalistoja täytyi siirtää transkriptoitavaksi, tehdyt transkriptiot täytyi tarkastaa ja hallinnoijien täytyi vastata käyttäjien kyselyihin. Havaittiin, että kirjaston on hyvä ilmoittaa sivustolla, että kirjasto omistaa oikeudet joukkoistamalla tuotettuun dataan ja transkriptioon.

Kehitysehdotuksena kirjoittajat esittivät käyttäjien motivoimiseksi mielenkiintoisen aiheiston valintaa. Uniikit ja paikalliset aiheet ovat hyviä, mutta toisaalta myös universaalit aiheet kuten ruoka ja sukututkimus kiinnostavat käyttäjiä. Käyttäjiä on sitoutettava tunnetasolla, esimerkiksi vetoamalla osallistumisen sijaan käyttäjää auttamaan kirjastoa. Joukkoistamishankkeen tavoitteen on oltava selkeä, erityisesti se mihin käyttäjiä tarvitaan. Tehtävät olisi pilkottava mahdollisimman pieniksi, jopa sana kerrallaan tehtäväksi. Palautetta käyttäjille voidaan antaa kiittämällä ja kannustamalla ja näytetään julkisesti, miten joukkoistamisella tehty työ etenee. Joukkoistamishankkeessa kannattaa luoda yhteisö, jonka palautetta kuunnellaan.

(Vershbow 2013; Lascarides & Vershbow 2014.)

4.1.2 Database of the Smokies

Mark Patrick Baggett, Anne Bridges, Ken Wise ja Sarah Tanner (2014) kirjoittivat Tennesseen yliopiston kirjaston ”Database of the Smokies” - tietokannan (DOTS) joukkoistamishankkeesta, jossa käyttäjien tehtävänä oli tågityksen sekä viitteiden ja/tai kokotekstien lisääminen tietokantaan. Käytännössä käyttäjät eivät tehneet tehtäviä suoraan tietokantaan vaan selainpohjaiselle alustalle (Drupal), joka oli tehty avoimella lähdekoodilla. Käyttäjillä oli käytössä kontrolloitu asiasanasto, mutta käyttäjien tågeistä muodostui vapaamuotoisempi folksonomia.

Käyttäjiä pyydettiin mukaan suoraan sähköpostilla kirjaston aiempien kontaktitietojen kautta. Lisäksi joukkoistamishankkeesta oli artikkeli ja osallistumiskutsu Great Smoky

Mountains Regional Project - hankkeen uutiskirjeessä, paikallislehdissä ja blogeissa. Joukkoistamiseen osallistuvat käyttäjät rekisteröityivät ja perustivat käyttäjätilin, jota hallinnoijat pystyivät kontrolloimaan. Hallinnoijat pystyivät kommunikoidaan myös sähköpostin avulla käyttäjän kanssa.

Laadun kontrolli rakennettiin prosessiin siten, että käyttäjät eivät olleet suoraan yhteydessä tietokantaan, vaan täyttivät lomakkeen. Lomakkeita oli neljä erilaista aineistotyypistä riippuen: artikkelille, kirjalle sekä muille dokumenteille kuten valokuville, viranomaisraporteille ja opinnäytteille. Neljäs lomakemalli oli avoin, siihen pystyi kopioimaan viitetiedot sellaisenaan. Kaikissa lomakkeissa oli mahdollisuus liittää mukaan kokoteksti. Lomakkeeseen käyttäjät lisäsivät aiheotsikon tai tegin. Hallinnoijat tarkastivat käyttäjien tekemät lomakkeet ennen niiden julkaisemista DOTS-tietokannassa.

Hankkeen neljän ensimmäisen kuukauden aikaiset 422 lokitiedostoa tai käyttäjän tehtävää analysoitiin tutkimalla viitteen oikeellisuus ja kattavuus sekä uudet aiheaikeukset taksonomiaan. Tarkoituksena oli selvittää, oliko hallinnointi tarpeen tietokannan laadun säilyttämiseksi. Tietueissa oli korjattavaa runsaasti: 63 tietuetta, joista 38:ssa oli korjattavaa kuten kirjoitusvirheitä, puuttuvia tietoja, vääriä muotoja kirjailijan nimestä, taksonomiaan sopimattomia asiasanoja. Tämän perusteella hallinnointia pidettiin tarpeellisenä.

Käyttäjien tekemien tegin arvioitiin olevan subjektiivista kuvailua, joten niiden ei nähty korvaavan ammattilaisen asiasanoitusta kontrolloidulla sanastolla vaan toimivan lisänä. Tegin olivat usein hyödyllisiä vaikka ne eivät olleet taksonomiassa. Tegin käyttöön rohkaistiin siten, että hallinnoija neuvoi tarvittaessa ja siitä tuli osa prosessia. 73 % tegin oli hyödyllisiä hallinnoijille, käyttäjien tegin avulla hallinnoijat saivat apua sisällönkuvailuun.

Ylimääräisenä, odottamattomana hyötynä saatiin tietokantaan lisätietoa ainutlaatuisesta ja harvinaisesta aineistosta. Tietokannan avaaminen mahdollisti yhteistyön myös muiden tutkimusyhteisöjen kanssa ja toisten kokoelmien "joukkoistamisen" DOTS-tietokantaan. Joukkoistamishankkeesta kirjoitettiin lehtiin ja siten tietoisuus tietokannasta kasvoi.

Joukkoistamisen lomakkeiden luomiseen, käyttäjien houkuttelemiseen ja hallinnointiin käytetty aika kannatti kirjoittajien mielestä. Pienestä henkilökunnasta tuli osa isompaa käyttäjien yhteisöä, jotka lisäsivät arvokasta sisältöä tietokantaan. Kehitysehdotuksena hankkeesta jäi tágien ja folksonomian parantaminen. Havaittiin, että tágerej voidaan käyttää asiasanojen lisn, mutta ne eivt korvaa kontrolloidun asiasanaston termej.

Kyttjtileist kirjasto aikoi myhemmin muodostaa keskustelufoorumin Smoky Mountain -alueen tutkimukselle. Kyttjist moni keskeytti tyskentelyn hankkeelle ajanpuutteen vuoksi sek sen thden, ett kyttj oli jo lisnnyt tiedossansa olevat viitteet. Motiivointi nhtiin trkeksi, mutta motivoivia tekijit ei hankkeessa selvitetty. Jatkossa hallinnoijat aikoivat hankkia uusia kyttj uutiskirjeen kautta ja suoraan shkpostitse. (Baggett, Bridges, Wise & Tanner 2014.)

4.1.3 Bentham Papers Transcription Initiative (Transcribe Bentham)

Tim Causer ja Melissa Terres (2014) esittelivt ”Transcribe Bentham” -joukkoistamisprojekti, joka oli kolmivaiheinen alkaen vuonna 2010. Kyttjien tehtvn oli kirjoittaa University College London Library’s Special Collectionissa olevat Benthamin ksikirjoitukset kokonaisuudessaan mediawiki-alustalle, joka pohjautuu avoimelle lhdekoodille. Kyttjt kirjoittivat transkription TEI (text encoding initiative) -standardilla tehdyille XML -lomakkeelle. TEI – merkkikieli mahdollisti ksikirjoituksissa olevan Benthamin omien poistojen ja lisysten siirron transkriptioon, konvertoinnin toiseen formaattiin sek pitkaikaissilytyksen. Kyttjien ei tarvinnut osata TEI-merkkikielt, vaan se oli koodattu valmiiksi lomakkeelle.

Joukkoistamiseen oli avoin kutsu. Aluksi oli vain yksi vapaaehtoinen kyttj, joka teki transkriptioita snnllisesti, ja muut 349 rekisteritynt kyttj tekivt vain satunnaisesti. Lehtijutun jlkeen kyttjien mr lisntyi ja transkriptioiden mr tuplaantui. Causerin ja Terresin kirjoittaessa artikkelia hankkeessa oli 2934 rekisteritynt kyttj, mutta vain 13 % kyttjist oli tehnyt kokonaisen transkriptiotehtvn. Nist 13 %:sta noin 2/3 oli tehnyt vain yht ksikirjoitusta. Kvi ilmi, ett 17 kyttj oli tehnyt suurimman osan joukkoistetuista tehtvist. Joukkoistaminen-termin kytt Transcribe

Bentham –hankkeen yhteydessä koettiin harhaanjohtavaksi, koska tuhannet käyttäjät eivät tehneet pieniä tehtäviä.

Käyttäjäkyselyn mukaan motivaationa käyttäjien osallistumiseen oli heidän oma kiinnostuksensa Benthamiin ja hänen tuotantoonsa, joukkoistamiseen tai teknologiaan. Myös halu osallistua yhteisen hyvän tekemiseen motivoi. Esteenä osallistumiseen mainittiin ajanpuute, Benthamin vaikea käsiala ja sen tulkinta sekä lomakepohjan merkkikieli. Käyttöliittymää parannettiin palautteen mukaan, esimerkiksi esikatselu-näkymä helpotti transkription tekoa TEI-merkkikielen käyttöliittymään. Tällä pyrittiin välttämään myös virheet. Hankkeessa pääteltiin, että joukkoistetun tehtävän pitäisi olla mahdollisimman helppo.

Hankkeen saatua lisärahoitusta mukaan tuli British Library, jolla myös on kokoelmissaan Benthamin käsikirjoituksia. Transkriptio-työkalua parannettiin automatisoimalla hallinnollisia prosesseja, käyttöliittymän näyttöä ja käsikirjoituksen valintaa parannettiin sekä tilastointia automatisoitiin. Esikatselu ja muutosten näkyvyys kertoivat käyttäjälle TEI-merkkikielen toiminnasta, ja niiden tavoitteena oli vähentää koodauskieleen liittyviä virheitä ja lisätä laatukontrollin tehokkuutta. Käytettävyyden parantamisen tavoitteena oli myös helpottaa rekrytointia.

Jatkoprojektin tranScriptorium, joka hyödynsi uutta Handwritten Text Recognition (HTR) (käsikirjoitetun tekstin tunnistus, vrt. koneella kirjoitetun tekstin optinen tekstin tunnistus eli OCR-teknologia). Riskinä oli, että käyttäjät kokevat taitojensa vähäteltävän, jos he uskovat koneen voivan korvata heidät. Joten hankkeessa pääteltiin HTR:n voivan täydentää käyttäjien työtä, ja käyttäjät voisivat kilpailla koneen kanssa tarkkuudesta.

Käyttäjä lähetti tehdyn lomakkeen hankkeen hallinnoijalle, joka tarkasti sekä tekstin että koodin. Jos virheitä oli vähän, työ lukittiin käyttäjälle korjattavaksi, mutta jos virheitä oli paljon tai työ oli jäänyt kesken, työ palautettiin joukkoistamisen tehtävälisälle. Lukitseminen toimi samalla laadun kontrollina käyttäjille. 95 % tehdyistä transkriptioista oli tutkijoiden mukaan korkealaatuista. 40 % transkriptioista hyväksyttiin ilman muutoksia. 14 % tarvitsi enemmän kuin 8 muutosta. Keskimääräisesti oli 4 korjausta per transkriptio.

Enemmän aiheutti korjauksia TEI-merkkikieli, jossa oli keskimäärin 8 korjausta per tehtävä. Vain 24 % transkriptioista oli täysin oikein. Eniten transkriptioita tehneet käyttäjät tarvitsivat keskimäärin 4 korjausta tekstiin ja 7 koodiin.

Hankkeessa kului yllättävän paljon resursseja ja henkilöstölle tuli uusia työtehtäviä kuten alustan kehittäminen ja ylläpito, käyttäjien rekrytointi ja hallinta, transkriptioiden tarkastus ja käsikirjoitusten asetusten standardointi. TEI-merkkikielen korjaus vaati enemmän työtä. Heräsi kysymys, olisivatko palkatut ammattilaiset tehneet työn tehokkaammin verrattuna joukkoistamisen hallinnointiin kuluneisiin resursseihin. Tätä pystyttiin vertailemaan, sillä vuosina 1984-2010 transkriptioita tehneet ammattilaiset tekivät vuodessa keskimäärin 1076 transkriptiota. Joukkoistamishankkeessa vuosina 2010-2013 vapaaehtoiset käyttäjät tekivät keskimäärin 2024 transkriptiota vuodessa. Joukkoistaminen nopeutti huomattavasti käsikirjoitusten transkriptioiden tekoa. Hankkeen arvioitiin säästäneen n. miljoona puntaa henkilöstökuluja. Vapaaehtoiset käyttäjät kehittyivät työssään, heille tuli sekä nopeutta että tarkkuutta lisää. Myös hallinnoijien laaduntarkastus nopeutui. Lisäksi joukkoistaminen toi julkisuutta ja mainosta käsikirjoituskokoelmalle. Ilman hallinnointia ja laaduntarkkailua tulokset olisivat kuitenkin kyseenalaisia.

Suurin osa transkriptioista tehtiin muutaman käyttäjän toimesta. Käyttäjäkyselyn mukaan syynä oli annetun tehtävän vaikeus. Tehtävät eivät olleet pieniä tehtäviä vaan laajoja, kokonaisia käsikirjoituksia, joiden pituus ja vaikeus vaihtelivat. Kehitysehdotuksena oli löytää ”superkäyttäjiä” ja pilkkoa kokonaisia, laajoja käsikirjoituksia pienemmiksi tehtäviksi.

(Causer & Terras 2014.)

4.1.4 Papers of the War Department 1784-1800

Sharon M. Leon (2014) esitteli ”Papers of the War Department 1784-1800” (PWD) hankkeen, jossa digitoitujen käsikirjoitettujen asiakirja-aineistojen transkriptiot tehtiin joukkoistamalla. Alkuperäiset asiakirjat olivat tuhoutuneet tulipalossa vuonna 1800. PWD:llä itsellään ei ollut alkuperäisiä asiakirjoita vaan niitä alettiin kerätä vuonna 1989 ympäri maailmaa muista arkistoista, ensin paperikopioina ja sitten sähköisessä muodossa. Vuonna 2007 avattiin nettisivu, jolla oli 45 000 dokumenttia. Ammattilaiset indeksoivat

ne perusmetadatalta: tekijä, vastaanottaja ja päivämäärä. Tämä ei ollut riittävä metadata tiedonhakuja ajatellen. Koska PWD:n resurssit eivät olleet riittävät itse tekemään lisää metadataa, he avasivat avoimen lähdekoodin työkalun transkriptioihin ja niiden tietoihin. Hankkeessa kehitettiin avoimen lähdekoodin työkalu, Scripto. Transkriptiot tehtiin avoimella lähdekoodilla mediawikin alustalle. Koska käyttäjä sai valita tehtävän, niin joukkoistaminen toimi samalla mittarina sille mitä aineistoa kysyttiin. Digitoitu käsikirjoitus ja transkriptio olivat samassa näkymässä, yllä oli alkuperäinen teksti ja alla transkriptio-tila, joka mahdollisti kokonaisen tekstirivin näkyvyyden.

Joukkoistamiseen osallistuneita käyttäjiä motivoi kiinnostus Amerikan historiaan, kansalaisvelvollisuuden tunto, tutkimuksen kohde, sukututkimus, koulutehtävät ja uteliaisuus transkription tekoon. 8 % käyttäjistä oli kirjasto- ja arkistoalan ihmisiä, jotka olivat ammatillisesti kiinnostuneita transkriptiotyökalusta. Hankkeen hallinnoijat kokivat käyttäjien sitouttamisen joukkoistamiseen tärkeäksi.

Transkriptioiden teko vaati rekisteröitymisen yhteydenpitoa varten ja syyn, miksi halusi osallistua joukkoistamiseen. Koska hallinnoijan kontrolli haluttiin säilyttää, mediawikiin ei sallittu anonyymien työtä. Käyttäjillä oli käytössä keskustelulomake, jossa käyttäjät voivat kysyä ja hallinnoijat vastata. Hallinnoijan tehtävänä oli myös blogata ja twiitata hankkeen edistymisestä.

Myös tässä joukkoistamishankkeessa maininta New York Timesissä toi lisää käyttäjiä. Joukkoistamiseen osallistui 1345 rekisteröitynyttä käyttäjää. Aktiivisia käyttäjiä oli 227, mutta vain 17 käyttäjää oli tehnyt enemmän kuin 100 dokumenttia. Dokumenttien määriä oli vaikea arvioida, tehtävän laajuus vaihteli muutaman rivin tekstistä monisatasivuiseen kirjaan. Kun joukkoistamisessa oli mukana 3800 käyttäjää, heidän jälkeensä tavoiteltiin uusia käyttäjiä blogilla, twitterillä, ja lähettämällä viestejä sähköpostilistoille. Kahden vuoden päästä oli jo 11 450 käyttäjää.

Käyttäjät pystyivät palaamaan aiempiin töihinsä ja korjaamaan niitä. Hallinnoijilla oli auktoriteetti suojella sivuja muiden editoinneilta ja siirtää transkriptiot mediawikistä Papers of War Department -tietokantaan.

(Leon 2014.)

4.1.5 Waisda?

Johan Oomen, Rite Gligorov ja Michiel Hildebrand (2014) esittelivät Waisda?-joukkoistamishankkeen, jossa käyttäjät tägittivät The Netherlands Institute for Sound and Vision –audiovisuaalisen arkiston videoita. Tägejä tarvittiin erilaisiin hakuihin: täsmälliseen tietoon, laajaan aihehakuun tai lyhyitä otoksia varten.

Tägitystä varten ulkopuolinen ohjelmistojen kehitysyhtiö kehitti pelin avoimella lähdekoodilla. Joukkoistaminen pelillistettiin siten, että antamalla tägin, joka on sama kuin toisen pelaajan jo antama, kumpikin sai pisteen. Pelissä kerättiin pisteitä sekä nopeudesta että tarkkuudesta. Videot pyörivät keskeytyksettä ja käyttäjien piti tägittää näkemäänsä. Pelin sekä arkiston kannalta tägi oli kelvollinen, kun kaksi pelaajaa kirjoitti saman tägin. Sitä kutsuttiin todennetuksi tägiksi. Käyttöliittymä antoi suoraa palautetta käyttäjälle, kuinka lähelle käyttäjän tägi osui jo hyväksytyjä tägejä. Käyttäjät näkivät kuinka he tägittivät suhteessa muihin pelaajiin. Jos läsnä ei ollut muita pelaajia, niin käyttäjät pelasivat aiempien pelien tägejä vastaan.

Waisda?-peliä pelasi 2000 käyttäjää, jotka antoivat 46 000 tägiä 600 videolle. 45,8 % pelaajista teki 1-10 tägiä. 35,3 % teki 10-100 tägiä ja 16,2 % teki 100-1000 tägiä. Vain 2,7 % käyttäjistä teki enemmän kuin 1000 tägiä. Kirjoittajat päättelivät, että tämän mukaan kannattaisi etsiä ”supertägitäjiä” suuren joukon sijasta. Käyttäjiä motivoi toiminta yhteisen hyvän puolesta, mutta videon sisältö motivoi myös. Suurin osa käyttäjistä halusi pelata samassa sessiossa toisia vastaan eikä aiemmin annettuja tägejä tavoitellen.

Tägien suhdetta sanastoihin tutkittiin ohjelmistosovelluksella. Vain 8 % tägeistä löytyi arkiston käyttämästä tesauruksesta, 23 % tägeistä löytyi hollannin kielen sanakirjasta. Kuilu näiden tägien välillä oli suurempi kuin todennetuilla tägeillä. 44 % todennetuista

tägeistä oli sanakirjassa, kun taas vain 14 % löytyi tesauruksesta. 9 % tägeistä löytyi molemmista. Todennetut tägit olivat kelvollisempaa hollantia kuin ei-todennetut. 59 % tägeistä ei löytynyt kummastakaan eikä niitä oltu todennettu. Melkein puolet näistä tägeistä koostui useammasta kuin yhdestä sanasta, mikä saattaa selittää sen ettei tägiä löytynyt sanastoista. Tägejä testattiin ottamalla lähempään tutkimukseen 200 tägiä, joihin ei ollut osunut yhtään hakua. Syiksi selvisi, että tägit olivat väärin kirjoitettuja tai täginä saattoi olla kokonainen lause, joka oli kielipillisesti virheellinen. Toinen virheryhmä tägejä oli sanat, joihin oli liitetty merkkejä, slangia, nimiä tai idiomia. Nimistä kirjoitettiin esimerkiksi vain sukunimi. Nämä esimerkit osoittivat kirjoittajille, että tägit eivät korvaa vaan täydentävät ammattilaisen työtä.

Tägien hyödyllisyyttä arkistolle tutkittiin vertaamalla eniten tägejä keränneen TV-shown jakson 738 tägiä arkiston ammattilaisten luettelointisääntöihin. Ammattiluettelot luokittivat tägejä asteikolla hyödytön-hyödyllinen-hyvin hyödyllinen. Hyödylliset tägit kuvasivat videota eri tavalla kuin ammattilaisten tekemät, sillä käyttäjien tägit kuvasivat lyhyitä hetkiä, kun ammattilaiset kuvailivat asiasanoilla laajempaa kokonaisuutta. Yhden videon otantana käyttäjien kokonaisuutta kuvaavista tägeistä vain 10 % oli hyödyllisiä, ja parhaiten tägitetyn videon 20 käytetyimmistä tägistä yksikään ei kuvannut koko videota. Tämä soti sitä oletusta vastaan, että mitä useammin tiettyä tägiä käytetään, sitä hyödyllisempi tägi on arkistolle.

Hankkeen toisessa vaiheessa tehtiin parannuksia peliin. Pelaajille annettiin lisäpisteitä, jos tägit osuivat sanastoihin. Myös synonyymeistä pelaaja sai pisteen. Käyttöliittymään lisättiin tägipilvi ja käyttäjäprofiilia tarkennettiin sekä siihen sisällytettiin pelihistoria. Analysoitu TV-show sisälsi 11 109 otosta. 4484 videota sai ainakin yhden tägin.

Joukkoistamisen onnistumista arvioitiin sen mukaan, kuinka tägeistä oli hyötyä tiedonhaussa. Tiedonhaku onnistui paremmin tägeillä kuin muun metadatan avulla. Paras tulos oli näiden kahden yhdistelmä. Pelkästään todennettujen tägien avulla tehty haku antoi huonomman tuloksen kuin kaikkien tägien avulla tehty haku. Havaittiin, että todennetut tägit olivat liian konservatiivisia ja hyödyllisiä kuvailuja jäi pois. Pelin avulla mikrotehtävien joukkoistaminen osoittautui hyödylliseksi, sillä hankkeessa annettiin yli miljoona tägiä. ”Superkäyttäjät” tekivät suurimman osan tägeistä. Useimmat tägeistä olivat hyödyllisiä ja ne täydensivät ammattilaisten sisällönkuvailua.

Joukkoistamishanketta markkinoitiin sosiaalisen median avulla käyttäjille. Palkintoja jaettiin käyttäjille ja heille järjestettiin myös tapahtumia. Sosiaalisia tägejä integroitiin käyttöliittymään ja esim. värikoodeilla eroteltiin tägit alkuperäisistä asiasanoista. Kontrolloidun asiasanaston käyttöä kokeiltiin ennakoivana tekstinsyöttönä. Olennaista vastaavissa hankkeissa olisi löytää superkäyttäjät, sekä määrälliset että laadulliset. Hankkeessa selvitettiin mahdollisuutta laadullisen ”superkäyttäjän” kontrollin vähentämiseen: käyttäjän annettua tietty määrä todennettuja tägejä käyttäjä sai luottamusta ja hänen ei todennettuja tägejä ei kontrolloitu. Käyttäjien osallistumisen motivaationa oli altruismi, pelin pelaaminen sekä kiinnostus ohjelmien ajankohtaisiin aiheisiin. (Oomen, Gligorov & Hildebrand 2014.)

4.1.6 Your Painting Tagger

Kathryn Eccles ja Andrew Greg kirjoittivat Your Painting Tagger –joukkoistamishankkeesta, jossa käyttäjät tägittivät maalauksia National Virtual Collectionista.

Hankkeessa määriteltiin metadatalle vaatimuksia:

- vapaasti luodut tägit
- maalauksen genren määrittely
- ihmisten ja paikkojen nimet erilleen aiheesta
- rakenteellinen, hierarkkinen aihesanojen sarja
- valmistusvuosi
- taiteelliset tyylit ja liikkeet
- taiteilijaryhmä

Tavoitteena oli määritellä 200 000 taululle tägit joukkoistamisen keinoin. Sosiaalisella tägityksellä tavoiteltiin myös uusia termejä, joita museoilla ei ollut vielä käytössä.

Rekisteröinnin oletettiin eliminoivan ilkeiden, jonka tosin tiedettiin olevan harvinaista näissä hankkeissa. Mutta rekisteröinti ei poistanut kirjoitusvirheitä eikä väärinymmärryksiä. Vaikka hankkeessa ohjeistettiin sanojen yksikkömuotoon, niin virheelliset muodot hyväksyttiin mieluummin kuin menetettiin hyviä tägejä.

Hankkeessa rakennettiin hierarkia kahdelle käyttäjäryhmälle, tavallisille ja asiantuntijoille tai muuten edistyneille. Asiantuntijoille, joita oli 7 % kaikista tägittäjistä, kohdentui maalauksen ajoitus sekä taiteilijan tyyli ja ryhmitys. Kaikille avoimen tägityksen katsottiin voitavan tehdä ilman taidehistorian tuntemusta ja heijastavan tavallisen käyttäjän tarpeita. Mukaan haettiin suurta yleisöä, koska aineistojen sisällönkuvailuun haluttiin käyttäjän näkökulma.

Tägit tarkastettiin automaattisesti Oxford English Dictionarystä (OED), mutta käyttäjät pystyivät kirjoittamaan OED:n päälle. Nimet tarkastettiin automaattisesti DBpediasta, joka mahdollisti myös linkit wikipediaan. Käyttäjien avuksi luotiin listoja, mutta taidehistorian tuntemus osoittautui tärkeäksi käyttäjille. Tägien laatua ei voitu kontrolloida täysin automaattisesti, joten hankkeen hallinnoijat valvoivat laatua.. Käyttäjällä oli mahdollisuus online-tutoriin.

Tägin hyväksymiseksi oli erilaisia vaatimuksia: lähtökohtaisesti vaadittiin, että ainakin kaksi käyttäjää antoi saman tägin, lajimäärittelyyn vaadittiin neljältä käyttäjältä sama tägi ja maalauksen ajoituksen määrittelyyn vaadittiin jopa kymmenen asiantuntijan tägi. Ongelmaksi tuli, että järjestelmä ei sallinut käyttäjän muuttaa antamaansa tägiä. Tähän kehitettiin ratkaisuksi se, että käyttäjät antoivat ensin tägejä näkemästään ja vasta sen jälkeen oli mahdollista määritellä yksityiskohtaisempaa metadataa listoja apuna käyttäen.

Hankkeessa koettiin ongelmaksi se, että mielenkiintoisimmiksi koetut maalaukset tägitettiin ensin ja runsaimmin. Hallinnoijat kokivat ongelmaksi tägien näkyvyyden tehtävässä, sillä se aiheutti lauma-ilmion, jossa näkyvä tägi johdatti ajattelua eikä ajattelu ollut itenäistä. Myös laatukontrollin automatisoinnin puute koettiin ongelmaksi. Ongelmien välttämiseksi kehitettiin käyttöliittymää siten, että käyttäjä ei voinut valita maalauksia eikä samaa maalausta voinut tehdä kahta kertaa. Maalaus katsottiin tägitetyn kattavasti kun 15 käyttäjää oli tägittänyt sen.

Motivaationa käyttäjillä oli kiinnostus maalauksiin, halu auttaa yleensä ja osallistua tutkimukseen. Maalauskoelma kiinnosti sinänsä sekä sen hyödyntäminen opetuksessa ja omassa tutkimuksessa. Haettiin myös oppia joukkoistamisesta. Osa käyttäjistä kaipasi so-

siaalista toimintaa kuten palautetta ja toimintaa hallinnoijien kanssa, joten käyttäjille järjestettiin tapaaminen hallinnoijien kanssa, jossa he saivat kuulla työnsä merkityksellisyydestä. Tämän jälkeen käyttäjät sitoutuivatkin joukkoistamiseen ja ehdottivat parannuksia käyttöliittymään. Puolet käyttäjistä halusi sosiaalisuutta ja yhteyden toisiin käyttäjiin.

Mukana joukkoistamisessa oli 9589 käyttäjää, joista 707 asiantuntijaa. He tägittivät 23 000 maalausta, joille annettiin yhteensä 4 miljoonaa tägiä. Käyttäjien tuottavuus vaihteli suuresti: 44 % käyttäjistä oli tägittänyt vähemmän kuin 5 maalausta. 65 käyttäjää oli tägittänyt yli 1000 maalausta. Joukkoistaminen oli kustannustehokasta ja työn jälki laadukasta. Omistautuneet käyttäjät tekevät säännöllisesti korkeatasoista työtä ja supertägittäjien merkitys oli suuri hankkeen toteutuksessa. Hallinnoijien mukaan yhteisöllisyyden tarve oli tärkeä havaita, samoin hankkeen kehittäminen käyttäjien palautteen mukaisesti. (Eccles &35 Greg 2014.)

5 TUTKIMUSTULOKSET

Opinnäytteessä tutkittiin, miten muistiorganisaatioiden aineiston transkriptio ja sisällönkuvailu voidaan organisoida suuren yleisön tehtäväksi joukkoistamisen avulla. Määriteltujen teemojen mukaan koottiin taulukot, joissa olevat numerot viittaavat luvussa 4 vastaavassa järjestyksessä esiteltyihin hankkeisiin. Eniten toistuvat havainnot on esitetty taulukossa ensimmäisenä.

	Toteutus	Havainnot
Tehtävä	Tägitys (2,5,6). Transkriptio (1,3,4) Metadatan (1) tai viitteiden ja aineiston (2) lisääminen.	Instituution on ilmoitettava sivustollaan, että se omistaa oikeudet joukkoistamalla tuotettuun dataan (1). Riskinä tekijänoikeudellisen aineiston lisääminen, sen kaltaisen aineiston lisääminen hallinnoijan kautta (2). Käyttäjän valitsema aineisto toimi myös mittarina kysynnälle (4). Ongelmaksi koettiin, että kysytyin aineisto tehtiin ensin (6).
Joukon rekrytointi	Käyttäjät rekisteröitiin yhteydenpitoa ja kontrollia varten. (2,3,4,5,6). Kohderyhmänä suuri yleisö avoimella kutsulla (1,3,4,5). Kohderyhmänä asiantuntijat tai alan harrastajat (2) Kohderyhmänä sekä suuri yleisö että asiantuntijat (6).	Lehtijuttujen jälkeen runsaasti uusia käyttäjiä. (1,2,3,4). Käyttäjien tuottavuus vaihteli: suurimman osan työstä teki muutama aktiivinen käyttäjä (3,4,5,6), jotka oppivat työn myötä nopeammiksi ja tarkemmiksi (3). Tehtävän käytettävyyttä parannettiin rekrytoinnin helpottamiseksi (3). Käyttäjien rekisteristä muodostettiin alan keskustelufoorumi (2). Käyttäjiä ei rekisteröity, jotta osallistumisen kynnyksen olisi matala (1).

Taulukko 1. Tehtävä ja rekrytointi

Tutkimusaineiston joukkoistamistehtävänä olivat tägitys tai transkriptio. Rekrytointi kohdentui neljässä esitellystä hankkeesta suurelle yleisölle avoimella kutsulla ja kahdessa hankkeessa rekrytointi kohdentui osittain suurelle yleisölle ja osittain asiantuntijoille. Asiantuntijoille kohdennettiin sisällönkuvailuun liittyvät joukkoistamiset. Käyttäjät rekisteröitiin hankkeessa sekä yhteydenpitoa että kontrollia varten. Joukkoistamisen uutisointi toi uusia käyttäjiä mukaan merkittävästi. Käyttäjien tuottavuus vaihteli suuresti,

mutta useamman hankkeen havainto oli, että suurimman osan tehtävistä teki pieni ”superkäyttäjien” joukko.

	Toteutus	Havainnot
Organisaation motivaatio	Haluttiin käyttäjän näkökulma sisällönkuvailuun (5,6) Haluttiin saada uusia termejä sisällönkuvailuun. (6)	Lehtijutuista markkinointihyötyä, koska tietoisuus aineistosta lisääntyi (2,3,6). Lisätietoa harvinaisesta aineistosta (2). Tietokannan avaamisen myötä mahdollistui yhteistyö muiden tutkimusyhteisöjen kanssa (2). Hallinnon resursseja kului enemmän kuin arvioitiin, silti kannattava hanke (3). Käyttäjien tekemä sisällönkuvailu ei korvaa ammattilaisen työtä, vaan täydentää sitä (5).
Käyttäjän motivaatio	Käyttäjiä motivoi oma kiinnostus aiheeseen (1,2,3,4,5,6) tai joukoistamiseen ja sen käyttämään teknologiaan (3,4, 6). Motivaationa oli myös kansalaisvelvollisuuden tunto ja halu osallistua yhteisen hyvän tekemiseen (3,4,5,6). Pelin pelaaminen (5).	Kiinnostusta herättää uniikit ja paikalliset aiheet, mutta myös universaalit kuten ruoka ja sukututkimus (1).
Käyttäjän motivointi	Käyttäjiä palkittiin ja heille järjestettiin yhteisöllisiä tapahtumia (5,6). Tehtävä jätettiin kesken, koska sitä pidettiin vaikeana vanhan käsialan ja tehtävän laajuuden tähden (3) tai ajanpuutteen vuoksi (2, 3). Lomakepohja koettiin hankalaksi (3) Käyttäjiä motivoitiin tehtävän peilistämällä (5).	Luodaan yhteisö, jota kuunnellaan (1,2,5,6). Tehtävä pilkottava pieniksi osiksi (1,3,5) tai tehtävän on oltava helppo (3). Tavoitteen oltava selkeä ja työn edistymisestä raportoitava julkisesti (1, 6) Palautetta käyttäjille on annettava kiittämällä ja palkitsemalla (1,5). Motivointi nähtiin tärkeäksi, mutta motivoivia tekijöitä ei selvitetty (2,4). Käyttäjät kilpailivat koneen kanssa tarkkuudesta (3). Käyttäjiä sitoutettava tunnetasolla, vetoamalla avuntarpeeseen (1).

Taulukko 2. Motivaatiot

Opinnäytteen tutkimustulosten mukaan käyttäjiä motivoi kiinnostava aihe ja myös halu osallistua yhteisen hyvän tekemiseen. Esitellyissä hankkeissa käyttäjät halusivat olla osa yhteisöä, joka teki töitä yhdessä. Hankkeissa motivaatiota pyrittiin lisäämällä tehtäviä ja kehittämällä ohjeistusta sekä palkitsemalla käyttäjiä. Viitekehyksessä esiteltyjen motiivointien lisäksi esille nousi pelillä motiivointi. Koska käyttäjien taso vaihteli, niin osalle tehtävät olivat liian vaikeita. Osa käyttäjistä kehittyi tarkemmaksi ja nopeammaksi harjaantuessaan. Tietynlaiset tehtävät vaativat asiantuntijuutta ja osalle käyttäjistä haasteet sopivat, mutta isolle osalle tehtävien olisi pitänyt olla helppoja ja pieniksi palasteltuna.

Organisaatioita motivoi joukkoistamisen aloittamiseen työvoiman saamisen lisäksi julkisuus, mutta hyötyjä ilmeni laajemminkin: organisaatio sai esimerkiksi lisätietoa harvinaisesta aineistosta ja yhteistyö lisääntyi yleisön sekä muiden asiantuntijayhteisöjen kanssa. Käyttäjien tэгitystyön ei nähty korvaavan ammattilaisen työtä vaan täydentävän sitä uusilla näkökulmilla eli semanttista kuilua saatiin siten kavennettua. Aineiston näkyvyys ja löydettävyys parantuivat.

	Toteutus	Havainnot
Työn laatu ja sen kontrolli	Käyttäjät tekivät tehtävän lomakkeelle tai alustalle (1,2,3,4,5,6). Käyttäjillä oli mahdollisuus kysyä neuvoa ja/tai antaa palautetta (1, 2, 3, 4, 5, 6). Hallinnoija tarkasti käyttäjän tekemän työn ennen julkaisua (1,2,3,4,6). Käyttäjillä oli apuna sisällönkuvailussa sanalistoja (2, 5, 6). Kontrollina toimi peli, kahden käyttäjän piti antaa sama tэгi (5,6) tai käyttäjät kilpailivat tarkkuudesta koneen kanssa (3). Käyttäjät pystyivät esikatselun avulla kontrolloimaan omia tai toisten töitä (1, 3). Työ lukittiin käyttäjälle korjausta varten (3).	Työn jälki oli korkealaatuista eikä väärinkäytöksiä havaittu (1,3,6). Tietueissa oli kirjoitusvirheitä, slangi-ilmaisuja, vajavaisia nimiä, lauseita. Myös asiantuntijoilla. (2,5,6). Virheelliset muodot hyväksyttiin (5,6). Laatuun vaikutti myös tekijän kokemus: mitä enemmän teki, sen parempi laatu (3). Laadukas työ toi mainetta tekijälle ja vähensi kontrollia (5). Ongelmaksi laumailmiö: näkyvät tэгit johdattelivat käyttäjän antamaan samoja tэгejä (6).

Taulukko 3. Työn laatu ja sen kontrolli

Joukkoistamisen avulla tehdyn työn laatua voitiin kontrolloida ohjeistuksella, suodattamalla tai korjaamalla virheellistä tietoa. Kaikissa muissa joukkoistamishankkeissa paitsi

yhdessä hallinnoijat tarkastivat käyttäjien tekemän työn ennen julkaisua. Hankkeessa 5 käyttäjät tägittivät pelin avulla ja virheettömästi tietyn määrän tehtäviä tehnyt käyttäjä saavutti luottamuksen, jolloin kontrollia vähennettiin hänen osaltaan. Käyttäjillä oli saalistat apuna useimmissa sisällönkuvailun joukkoistamishankkeissa.

	Toteutus	Havainnot
Roolit	Hallinnoija kontrolloi käyttäjän työtä (1,2,3,4,6). Hallinnoija viestitti, bloggasi ja tviit-tasi edistymisestä (1,4). Värikoodeilla erotettiin käyttäjän tägit ammatilaisen antamista asia-sanoista (5).	Käyttäjän tägit olivat hallinnoijan apuna sisällönkuvailussa. Tägit olivat subjektiivista kuvailua eivätkä korvanneet vaan täydensivät ammatilaisen asiansanoitusta (2,5). Joukkoistaminen kaksinkertaisti transkriptioiden määrän, mutta ilman hallinnointia ja laaduntarkkailua tulokset olisivat kyseenalaisia (3). Käyttäjiltä saatiin tietoa harvinaisesta aineistosta (2). Käyttäjän asiantuntemus osoittautui tärkeäksi (6). Joukkoistamisessa avautui organisatiolle yhteistyö muiden tutkimusyhteisöjen kanssa (2).

Taulukko 4. Hallinnoijan ja käyttäjän roolit

Esitellyissä hankkeissa ammatilaiset eivät luovuttaneet auktoriteettia käyttäjille. Joukkoistamishankkeissa ammatilaisen rooli oli toimia suunnittelijana ja käyttäjien toteuttajina. Yhdessäkään esitellyistä hankkeista ei tehty varsinaisesti yhteistyötä eivätkä käyttäjät saaneet vapaasti työstää aineistoa. Yhdessä hankkeessa (5) käyttäjän tägit eroteltiin värikoodein ammatilaisten asiansanoista. Hankkeessa 2, jossa oli mukana asiantuntijoita, käyttäjät antoivat tietoa harvinaisesta aineistosta. Myös hankkeessa 6 käyttäjien asiantuntemus osoittautui tärkeäksi.

6 JOHTOPÄÄTÖKSET

Arkistojen, kirjastojen ja museoiden tehtävät heijastelevat ympäröivää yhteiskuntaa ja siten yhteiskunnalliset muutokset muuttavat myös niiden tehtävää ja asemaa. Arkistojen rooli on muuttumassa säilyttävästä ja väärinkäytön estävästä varastosta kaikille avoimeksi tilaksi ja edelleen avoimeksi toiminnaksi: Passiivisesta säilyttäjistä aktiiviseksi kulttuuriperinnön muodostajaksi ja yhteisön muistin säilyttäjäksi, mutta myös muistojen muokkaajaksi. Myös asiakkaan rooli on muuttumassa passiivisesta käyttäjästä aktiiviseksi sisältöjen tuottajaksi. Muistiorganisaatioiden aineiston joukkoistamista ei ole vielä paljon tehty ja sitä on tutkittu vähän.

Tutkielman tiedonhaun haasteena oli artikkeleiden asiasanoituksen tason vaihtelu sekä laadullisesti että määrällisesti. Lisähaastetta tiedonhakuun toi tieteen alan vakiintumatonta terminologia kuten myös tietokantojen erilaiset hakumahdollisuudet. Tutkimusmenetelmäksi valitulla systemaattisella kirjallisuuskatsauksella pyrittiin kattavaan tiedonhakuun ja vastaamaan asiasanoituksen haasteisiin.

Tutkielman opinnäytteen kysymyksenasettelu rajasi haetun aineiston suppeaksi, koska etsittiin tietyillä teemoilla analysoituja joukkoistamishankkeita. Systemaattisen tiedonhaun tuloksena saatu aineisto kohdentui pääosin yhden monografian, Mia Ridgen toimittaman julkaisun ”Crowdsourcing our cultural heritage” artikkeleihin. Sitä selittänee se, että teoksen aiheet ja kysymyksenasettelu olivat yhteensopivia tämän tutkielman vastaaviin. Tosin em. teoksen artikkelit eivät käsitelleet suoraan rooleja tai auktoriteettia, jotka olivat tulkittavissa laadun kontrollin kautta. Ehkä joukkoistamishankkeita ei oltu vielä analysoitu tieteellisissä julkaisuissa riittävästi laadullisesti ja määrällisesti opinnäytteen kysymyksenasettelun edellyttämällä tavalla. Tutkielman tiedonhakua olisi voitu laajentaa koskemaan esimerkiksi blogeja tai artikkeleita konferenssijulkaisuissa tai tutkielman kysymyksenasettelua olisi voitu määritellä toisin. Tai olisi voitu valita toinen tutkimusmenetelmä kuin systemaattinen kirjallisuuskatsaus. Joukkoistamishankkeiden hallinnoijille kohdennetulla kysymyslomakkeella olisi voitu päästä kattavampaan, tarkempaan ja vertailukelpoisempaan tulokseen haluttujen teemojen tutkimisessa.

Tutkimustulokset vahvistivat aiemman Oomenin, Gligorovin ja Hildebrandin (2014) tutkimuksen tuloksia: organisaation motivaationa oli lisäresurssien saaminen tehtävien tekoon sekä käyttäjän näkökulman ja lisätiedon saaminen. Lisäksi joukkoistaminen toimi yhteydenpitokeinona käyttäjiin. Uutena tuloksena voisi mainita yhteistyön toisten organisaatioiden kanssa tietokantoja avattaessa sekä joukkoistamisen mukana tulleen julkisuuden joukkoistettavalle aineistolle ja organisaatiolle. Tosin julkisuuden suhteen on esitettävä pieni varaus: kyseessä voi olla joukkoistamiseen liittyvä uutuudenviehätys ja joukkoistamiselle avatun aineiston yleinen kiinnostavuuden taso.

Tutkielman tutkimustulokset tukivat myös Smith-Yoshimuran ym. laajan tutkimuksen suosituksia: selvien tavoitteiden määrittelyä sosiaalisen median käytössä, käyttäjien motivointia, avoimen lähdekoodin ohjelmistojen käyttöä, käyttäjien tekemän sisällön indeksointia ja hakukelpoisuutta joko erikseen tai yhdessä organisaation sisällön kanssa (Smith-Yoshimura & Holley 2012, 9-17).

Esitellyissä hankkeissa käyttäjät jaettiin suureen yleisöön ja asiantuntijoihin, jolloin rekrytointi voi olla avoimella tai rajatulla kutsulla. Tarvittaessa joukkoistamisen voitiin kohdentaa joko osittain tai kokonaan asiantuntijoille. Olennaista oli huomioida käyttäjien rekrytoinnissa tehtävä: onko tehtävä mekaaninen vai asiantuntemusta vaativa. Tehtävän laadun lisäksi piti huomioida sen laajuus. Tehtävä ei saanut olla liian helppo eikä liian vaikea, mutta ei myöskään liian laaja, jotta käyttäjän motivaatio säilyi. Olennaista motivaatiolle oli käyttäjän kiinnostus aiheeseen, kuten Oomen, Gligorov ja Hildebrand (2014) sekä Huvila (2015) ja Holley (2010) olivat omissa aiemmissä tutkimuksissaan todenneet. Tutkimustulokset vahvistivat viitekehyksessä esitellyn Oomen ja Aroyonin (2011) sekä Holleyn (2010) näkemyksen, jonka mukaan käyttäjät halusivat olla osa yhteisöä. Tämän tutkielman yhdessä esitellyssä hankkeessa esitettiin motivointina tehtävän haastavuuden lisäämistä työn edetessä, koska käyttäjät kehittyivät työssään. Tällaista motivointia ei kuitenkaan oltu toteutettu esitellyissä hankkeissa.

Auktoriteettia ei esitellyissä hankkeissa tuotu terminä esille, vaan se ilmeni hallinnoijan ja käyttäjien rooleissa sekä käyttäjän työn laadun kontrollissa. Ammatillaisen rooli joukkoistamisessa ilmeni siinä, kuinka paljon hän teki yhteistyötä käyttäjien kanssa jakaen

omia työtehtäviään vai kontrolloiko hän vapaaehtoisia. Hallinnoijat ja käyttäjät eivät toiminteet esitellyissä hankkeissa tasavertaisia. Hallinnoijat eivät luovuttaneet auktoriteettia käyttäjille, vaan hallinnoijat kontrolloivat käyttäjien työtä tarkastamalla heidän tekemänsä tehtävät ennen julkaisua. Tarkastamalla organisaatio auktorisoi tehdyn työn: tehty työ tarkastettiin ennen kuin organisaatio otti sen oman auktoriteetin alle julkaistavaksi. Tämä vahvistaa Yakelin (2011) sekä Oomenin ja Aroyon (2011) näkemystä arkistohoitajan roolista erillisenä kontrolloivana toimijana: tutkimusten mukaan laatua voitiin kontrolloida erottelemalla joukkoistamalla tehty työ ammattilaisen työstä, ohjeistamalla ja tarkastamalla joukkoistamalla tehty työ ennen julkaisua.

Joukkoistamisen onnistumiseen vaikutti sekä tekijöiden osaamisen taso että tehtävän laajuus ja sen vaikeusaste. Tehtävän onnistumiseen vaikutti myös se, oliko kyseessä transkription kaltainen mekaaninen toiminta vai tarvittiinko sisällönkuvailuun aiheentunteusta. Hallinnoijat pystyivät ennakkoon minimoimaan ongelmia ohjeilla, lomakepohjilla ja neuvonnalla, mutta ne veivät hallinnon resursseja.

Koska käyttäjien taso vaihteli, osalle käyttäjistä tehtävät olivat liian vaikeita. Tietynlaiset tehtävät vaativat asiantuntijuutta ja osalle käyttäjistä haasteet sopivat, mutta isolle osalle käyttäjiä tehtävien olisi pitänyt olla helppoja ja pieniksi palasteltuja. Paljon tekevät käyttäjät kehittyivät tarkemmiksi, jolloin työn laatu parani. Uutena tutkimustuloksena tuli näiden laadukasta työtä tekevien superkäyttäjien rekrytoiminen. Heidän rekrytoimiseensa esitellyillä hankkeilla ei ollut ratkaisua. Lehtijutuilla oli suuri merkitys rekrytoinneissa, mutta tulivatko superkäyttäjät lehtien kautta? Tulivatko he suuren joukon mukana vai kohdennetulla kutsulla? Nämä keskeiset seikat eivät selvinneet tutkimusaineistosta. Näiden laadullisten ja/tai määrällisten superkäyttäjien löytäminen olisi yksi jatkotutkimuksen aihe.

Joukkoistamisessa olisi syytä erottaa suurelle yleisölle suunnattavat mekaaniset tehtävät ja asiantuntijoille suunnattavat erikoisosaamista vaativat tehtävät. Organisaatioita motivoi joukkoistamiseen myös asiantuntijatiedon saaminen. Tämä osoitti asiantuntijoille kohdennetun joukkoistamisen, nichesourcingin, erikoislaadun, jolla oli vaikutusta rooleihin, työn laadun kontrolliin ja auktoriteetin jakamiseen. Tällöin ammattilaisen ja käyttäjän roolien jako olisi tasavertaisempi ja luottamuksellisempi. Yllättävää oli, että yhdessäkään joukkoistamishankkeen artikkelissa ei otettu esiin käsitettä *nichesourcing*, vaikka

asiantuntijoille kohdennettua joukkoistamista olikin kahdessa hankkeessa. Auktoriteettia voisi luovuttaa asiantuntijoille tai sellaisille suuren yleisön edustajille, jotka ovat saaneet luottamuksen osoittamalla päässeensä tietylle tasolle tehtävän annossa. Asiantuntijoille suunnatun joukkoistamisen tutkiminen, nichesourcing, ja siihen liittyvä auktoriteetin jako olisikin hyvä jatkotutkimuksen aihe.

Ristiriita ilmeni työn laadun tuloksissa: kahdessa transkriptiohankkeessa ja yhdessä sisällönkuvailuhankkeessa työn jälki oli hallinnoijien mielestä korkealaatuista, kun taas kahdessa sisällönkuvailuhankkeessa tietueissa oli runsaasti korjattavaa. Yhdessä transkriptiohankkeessa ei eritelty työn laatua. Lähemmässä tarkastelussa sisällönkuvailuhankkeiden sisällönkuvailun laatutasolle oli eri määritelmät: hankkeessa 6 virheelliset muodot hyväksyttiin, jotta ei menetettäisi kenties arvokasta tietoa. Mutta hankkeessa 5 havaittiin ongelmaksi se, että näitä hyviä, mutta väärinkirjoitettuja termejä ei löytynyt haussa. Nämä arviot työn jäljen laadusta eivät olleet yhteismitallisia, joten niitä ei voi verrata keskenään vaan on tyytyminen hallinnoijien omaan suhteelliseen arvioon omasta hankkeestaan. Huomionarvoista oli, että ilkeävaltaa ei ilmoitettu havaitun yhdessäkään hankkeessa, ei edes siinä hankkeessa, jossa ei vaadittu rekisteröitymistä. Ehkä näiden hyvien kokemusten jälkeen uskalletaan kokeilla auktoriteetin jakoa tulevissa joukkoistamishankkeissa.

Tutkimuksessa kävi ilmi, että aiempaan tutkimukseen nähden uutena nousi esiin pelien käyttö joukkoistamisessa. Pelillä voitiin tehdä joko itse tehtävä tai peli toimii laadun kontrollina tai motivaationa. Yksi jatkotutkimuksen aihe voisi olla pelien avulla tehdyn joukkoistamisen tutkimus.

Tutkimustuloksista ilmeni, että joukkoistamisen toteutukset ja havainnot olivat samansuuntaisia esitellyissä hankkeissa ja siten sovellettavissa yleisemmin. Joukkoistamisen organisoinnissa oli monta muuttujaa, joten yhdellä kaavalla ei voi toteuttaa kaikkia joukkoistamisia, mutta yhteneväisiä piirteitä löytyi. Opinnäytteessä esiteltyjen hankkeiden yksittäisiäkin huomioita voi soveltaa tulevien joukkoistamishankkeiden käytänteiksi. Jotta saadaan lisää käytänteitä, joukkoistamishankkeita on tehtävä enemmän.

7 LÄHTEET

Cook, Terry 2013: Evidence, memory, identity, and community: four shifting archival paradigms. *Archival Science* 13 (2-3), 95-120.

de Boer V. ym. 2012: Nichesourcing: Harnessing the power of crowds of experts. Teoksessa ten Teije A. ym. (toim.) *Knowledge Engineering and Knowledge Management*. EKAW 2012 Lecture Notes in Computer Science, 7603. Berlin, Heidelberg: Springer, 16-20.

Doan, Anhai, Ramakrishnan, Raghu, Halevy, Alon Y. 2011: Crowdsourcing systems on the World-Wide-Web. *Communications of the ACM* 54 (4), 86-96.

Estellés-Arolas, E. & González-Ladrón-de-Guevara, F. 2012: Towards an integrated crowdsourcing definition. *Journal of Information Science* 38 (2), 189-200.

Evans, Max J. 2007: Archives of the people, by the people, for the people. *The American Archivist* 70 (2), 387-400.

Fink, Arlene 2014: *Conducting research literature reviews*. From the internet to paper. Thousand Oaks: Sage.

Gorzalski, Matt 2013: Examining user-created description in the archival profession. *Journal of Archival Organization* 11 (1-2), 1-22.

Holley, Rose 2010: Crowdsourcing. How and why should libraries do it? *D-Lib Magazine* 16 (4), NP.

Hossain, Mokter & Kauranen, Ilkka 2015: Crowdsourcing: a comprehensive literature review. *Strategic Outsourcing: An International Journal* 8 (1), 2-22.

Howe, Jeff 2006: The rise of crowdsourcing. *Wired* 6 (14), 1-4.

Huvila, Isto 2015: The unbearable lightness of participating? Revisiting the discourses of “participation” in archival literature. *Journal of Documentation* 71 (2), 358-386.

Ketelaar, E. 1992: Archives of the people by the people, for the people. *S.A. Archives Journal* 34, 5-16.

Lehtiö, Leeni & Johansson, Elise 2016: Järjestelmällinen tiedonhaku hoitotieteessä. Teoksessa Stolt, Minna, Axelin, Anna, Suhonen, Riitta (toim.) *Kirjallisuuskatsaus hoitotieteessä*. Hoitotieteen laitoksen julkaisuja Tutkimuksia ja raportteja Sarja A 73. Turku: Turun yliopisto, 35-55.

Lemetti, Terhi & Ylönen, Minna 2016: Kirjallisuuskatsaukseen valittujen tutkimusartikkelien arviointi. Teoksessa Stolt, Minna, Axelin, Anna, Suhonen, Riitta (toim.) *Kirjallisuuskatsaus hoitotieteessä*. Hoitotieteen laitoksen julkaisuja Tutkimuksia ja raportteja Sarja A 73. Turku: Turun yliopisto, 67-79.

Li Liew, Chen, Wellington, Shannon, Oliver, Gillian & Perkins, Reid 2015: Social media in libraries and archives: Applied with caution. *The Canadian Journal of Information and Library Science* 39 (3-4), 377-396.

Niela-Vilén, Hannakaisa & Hamari, Lotta 2016: Kirjallisuuskatsauksen vaiheet. Teoksessa Stolt, Minna, Axelin, Anna, Suhonen, Riitta (toim.) *Kirjallisuuskatsaus hoitotieteessä*. Hoitotieteen laitoksen julkaisuja Tutkimuksia ja raportteja Sarja A 73. Turku: Turun yliopisto, 23-34.

Oomen, Johan & Aroyo, Lora 2011: Crowdsourcing in the cultural heritage domain: Opportunities and challenges. *C&T'11, 29 June - 2 July 2011, QUT, Brisbane, Australia*, 138-149.

Petek, Marija 2012: Comparing user-generated and librarian-generated metadata on digital images. *OCLC Systems & Services: International digital library perspectives*, 28 (2), 101-111.

Ridge, Mia (toim.) 2014: *Crowdsourcing our cultural heritage*. Farnham: Ashgate.

Salminen, Ari 2011: *Mikä kirjallisuuskatsaus?* Johdatus kirjallisuuskatsauksen tyyppeihin ja hallintotieteellisiin sovelluksiin. Vaasan yliopiston julkaisuja, opetusjulkaisuja 62, julkisjohtaminen 4. Vaasa: Vaasan yliopisto.

http://www.uva.fi/materiaali/pdf/isbn_978-952-476-349-3.pdf (käytetty 13.5.2017)

Smith-Yoshimura, Karen & Shein, Cyndi 2011: *Social metadata for libraries, archives and museums. Part 1: Site reviews*. Dublin, Ohio: OCLC.

<http://www.oclc.org/research/publications/library/2011/2011-02.pdf>

(käytetty 13.5.2017)

Smith-Yoshimura, Karen, Godby, Carol Jean, Koffler, Helice, Varnum, Ken & Yakel, Elizabeth 2011: *Social metadata for libraries, archives and museums. Part 2: Survey analysis*. Dublin, Ohio: OCLC Research.

<http://www.oclc.org/research/publications/library/2011/2011-03.pdf>

(käytetty 13.5.2017)

Smith-Yoshimura, Karen & Holley, Rose 2012: *Social metadata for libraries, archives, and museums. Part 3: Recommendations and readings*. Dublin, Ohio: OCLC Research.

<http://www.oclc.org/research/publications/library/2012/2012-01.pdf>

(käytetty 13.5.2017)

Suhonen, Riitta, Axelin, Anna & Stolt, Minna 2016: Erilaiset kirjallisuuskatsaukset. Teoksessa Stolt, Minna, Axelin, Anna, Suhonen, Riitta (toim.) *Kirjallisuuskatsaus hoitotieteessä*. Hoitotieteen laitoksen julkaisuja Tutkimuksia ja raportteja Sarja A 73. Turku: Turun yliopisto, 7-22.

Theimer, Kate 2011: What is the meaning of Archives 2.0? *The American Archivist* 74 (1), 58-68.

Tieteen termipankki. <http://tieteentermipankki.fi> (käytetty 13.5.2017)

Valkeapää, Kirsi 2016: Tutkimusaineiston valinta systemaattisessa kirjallisuuskatsauksessa. Teoksessa Stolt, Minna, Axelin, Anna, Suhonen, Riitta (toim.) *Kirjallisuuskatsaus*

hoitotieteessä. Hoitotieteen laitoksen julkaisuja Tutkimuksia ja raportteja Sarja A 73. Turku: Turun yliopisto, 56-66.

Yakel, Elizabeth 2011: Balancing archival authority with encouraging authentic voices to engage with records. Teoksessa Theimer, Kate (ed.) *A different kind of web*. New connections between archives and our users. Chicago: Society of American Archivist, 75-101.

LIITE: TUTKIMUSAINEISTONA KÄYTETYT ARTIKKELIT

Baggett, Mark Patrick, Bridges, Anne, Wise, Ken, Tanner Sarah, Mezick Jennifer 2014: Populating the wilderness: crowdsourcing database of the Smokies. *Library Hi Tech*, 32 (2), 249-259.

Causer, Tim, Terras, Melissa 2014: "Many hands make light work, many hands together merry work": Transcribe Bentham and crowdsourcing manuscript collections. Teoksessa Ridge, Mia (toim.): *Crowdsourcing our cultural heritage*. Farnham: Ashgate, 57-88.

Eccles, Kathryn, Greg, Andrew 2014: Your paintings tagger: Crowdsourcing descriptive metadata for a national virtual collection. Teoksessa Ridge, Mia (toim.): *Crowdsourcing our cultural heritage*. Farnham: Ashgate, 185-210.

Lascarides, Michael, Vershbow, Ben 2014: What's on the menu?: Crowdsourcing at the New York Public Library. Teoksessa Ridge, Mia (toim.): *Crowdsourcing our cultural heritage*. Farnham: Ashgate, 113-138.

Leon, Sharon M. 2014: Build, analyse and generalise. Community transcription of the Papers of the War Department and the development of Scripto. Teoksessa Ridge, Mia (toim.): *Crowdsourcing our cultural heritage*. Farnham: Ashgate, 89-112.

Oomen, Johan, Gligorov, Rite, Hildebrang, Michiel 2014: Waisda?: Making videos findable through crowdsourced annotations. Teoksessa Ridge, Mia (toim.): *Crowdsourcing our cultural heritage*. Farnham: Ashgate, 161-184.

Vershbow, Ben 2013: NYPL Labs: Hacking the library. *Journal of Library Administration*, 53 (1), 79-96.