



**Työterveyslaitos** | Arbetshälsöinstitutet  
Finnish Institute of Occupational Health

# Työdata ja tekoäly

TOIMINNAN KEHITTÄMISEN JA JOHTAMISEN TUEKSI

Tiina Kalliomäki-Levanto  
Jussi Korpela  
Erkki Sinisalo  
Riku Louhimo





**Työterveyslaitos** | Arbetshälsoinstitutet  
Finnish Institute of Occupational Health

## Työdata ja tekoäly

TOIMINNAN KEHITTÄMISEN JA JOHTAMISEN TUEKSI

Tiina Kalliomäki-Levanto

Jussi Korpela

Erkki Sinisalo

Riku Louhimo

Työterveyslaitos

Helsinki



Työterveyslaitos

PL 40

00032 Työterveyslaitos

[www.ttl.fi](http://www.ttl.fi)

Piirroksat: Vilja Rydman (kuva 4)

© 2019 Työterveyslaitos ja kirjoittajat

Julkaisu on toteutettu Työsuojelurahaston tuella.

Tämän teoksen osittainenkin kopiointi on tekijänoikeuslain (404/61, siihen myöhemmin tehtyine muutoksineen) mukaisesti kielletty ilman asianmukaista lupaa.

ISBN 978-952-261-909-9 (nid.)

ISBN 978-952-261-908-2 (PDF)

PunaMusta Oy, Tampere, 2019

## TIIVISTELMÄ

Tässä raportissa kerrotaan Työsuojelurahaston rahoittaman ”Työdatan hiljaiset signaalit” – TYHIS-hankkeen työvaiheista, joilla työelämä tutkimuksen uusi aineistolähde ”digital trace data”, työdata, saatiin käytettävään muotoon. Esittelemme analyysitavan ja periaatteet, joita voidaan soveltaa moniin vastaaviin aineistolähteisiin. Työdataa tarkastella koko organisaation yhteisen tekemisen näkökulmasta. Tutkimuksessa kehitettyä menetelmää voidaan suoraan hyödyntää työhyvinvoinnin kehittämiseen asiantuntijaorganisaatioissa.

Raportissa kerrotaan ensin toiminnallisen työdatan yleisiä piirteitä ja siitä, miten toiminnallinen työdata eroaa rakenteellisesta ”raportointi” -datasta. Sitten siirrytään hankkeen työdataan, joka on yhden ammattikorkeakoulun Moodlen lokitietoihin jäävää tietoa. Moodle on virtuaalinen oppimisympäristö, jota opettajat käyttävät opetuksen suunnitteluun, toteutukseen sekä opiskelijoiden edistymisen seurantaan ja arviointiin.

Tutkimusaineisto sisältää 2,8 miljoonaa käyttötapahtumaa 63 viikon jaksolta. Käyttötapahtumia on rajallinen määrä, yhteensä 181 kappaletta. Käyttötapahtumat ovat lokitiedossa sanapareina, esimerkiksi ”message\_sent”, ”submission\_graded” tai ”course\_module\_updated”. Käyttötapahtumia tarkastellaan tekstinä, joka ryhmitellään opettajakohtaisiksi viikon jaksoiksi. Näistä muodostuu 10 210 opeviikkotekstiä. Teksteissä sanaparien keskinäiset suhteet muodostavat jakaumia, esimerkiksi jonain viikkoina tehdään enemmän arviointia ja toisina kurssien suunnittelua.

Työdatan sisältämä tieto työn tekemisestä liitetään työn käytäntöjen viitekehykseen. Käytäntöjä sisältävää tekstiaineistoa analysoivana tekoälynä sovelletaan LDA nimistä aihealinnusta. LDA analysoi tekstien sisältämien sanojen jakaumia. LDA, käytännöt ja työdata jakavat yhteisen periaatteen havaintojen välisistä suhteista. Rajatussa aineiston osassa, dokumentissa, havainnot (sanat, käytännöt, sanaparit) muodostavat jakaumia.

Tekoäly tunnisti 17 käytäntöä, jotka voitiin karkeasti jakaa kolmeen ryhmään: 1) tietolähteiden ja opiskelijoiden edistymisen *seurantaan*, 2) *tekemiseen* sekä 3) käytäntöihin, joissa on *sekä seuranta että tekemistä*. Käytäntöjä toteutettiin erilaisin intensiteetein eri viikkoina.

Viikkotason tarkastelu tehtiin käytäntöjen lisäksi sairauspoissaolotiedoista ja suunnittelu-järjestelmistä saaduista opetustyöhön ja opettajien keskinäiseen yhteistyöhön liittyvistä tiedoista. Käytäntöjen ja muiden työhön liittyvien tekijöiden viikkorytmiä verrattiin poissaolorytmiin. Vertailu tehtiin visuaalisesti ja huippujen tunnistusmenetelmällä. Hankkeen keskeisin tulos on edellä kuvattu menetelmällinen prosessi, jossa eri lähteiden data muokataan viikkotason vertailut mahdollistavaksi välineeksi.



Lopuksi pohdittiin aikasarjatiedon ja huippujen tunnistamisen hyödyntämistä organisaatioissa ja tulevaisuuden mahdollisuuksia. Kun tulevaisuudessa rytmien seuranta voi olla automaattista, voi käytäntöjen kehittäminen ja kehittämisen tulosten seurantakin olla jatkuvaa. Vielä pohdittiin aineistolähdettä ja viikkorytmiä uutuuksina suhteessa olemassa oleviin menetelmiin ja malleihin tutkia työtä ja työhyvinvointia.

Etsimme myös tutkimuskirjallisuudesta "digital trace dataa", työdataa, hyödyntäviä artikkeleita. Lähteet -osassa 106 artikkelia on ryhmitelty seitsemään ryhmään. Löydetyt artikkelit olivat

- 1) yleisluontoisia big datan, trace datan ja analytiikan pohdintoja,
- 2) yleistä tutkimuksen metodologiasta kuten subjektiivinen vai objektiivinen tieto, datalähtöinen vai mallilähtöinen tutkimusote,
- 3) HR -datan hyödyntämistä pohtivia,
- 4) ehdotuksia datatieteilijöiden näkökulmasta työdatan hyödyntämiseksi,
- 5) organisaatiodynamiikkaa, prosesseja ja aikaa sekä digital trace dataa käsitteleviä artikkeleita, joita voisi soveltaa "työn aikasarja" tietoon,
- 6) asiantuntijan työaikaa organisaatiotason ilmiönä käsittelevät tutkimukset,
- 7) Moodle-dataa, "virtual learning environment" -dataa, hyödyntävät "learning analytics" teemaiset tutkimukset, joissa ennakoitiin *opiskelijan* edistymistä.

Lopussa, viimeisessä liitteessä, löytyy hankkeen aikana kirjoitetut blogit.



## ABSTRACT

This report describes the phases of the "Silent Signals of Work Data" -project funded by The Finnish Work Environment Fund. Through the phases, a new "digital trace data", the work data, was developed to the form of a tool that can be used for both research and development. We present a method of analysis and principles that can be applied to many similar sources of data. We interpret work data from an organizational perspective, not an individual.

We first outline the general features of functional work data and describe how functional work data differs from structured "reporting" data. Then we move on to the project's work data, which is the log data of Moodle from one university of applied science. Moodle is a virtual learning environment used by teachers to plan and implement teaching and to monitor and evaluate student progress.

The data contains 2,8 million rows of usage events over a 63-week period. Usage events in the log data are expressed in pairs of words, for example "message\_sent", "submission\_graded" or "course\_module\_updated". The data comprises of 181 different usage events, representing a limited variety of possible events.

Usage events are analyzed as text grouped into teacher-specific sections of the week. The total number of these weekly text-sections is 10 210. In one text-section, the relationships between the pairs of words (i.e. usage events) form distributions, for example, there is more evaluation in some weeks and course planning in others.

The information of working contained in the work data is attached to the framework of practice theory. Artificial intelligence analyzing textual material (10 210) applies LDA-themed modeling. The LDA analyzes the word distributions in the texts. LDA, practices and work data share a common principle of relationships between observations. In the document, observations (words, practices, word pairs) form distributions.

Artificial Intelligence identified 17 practices that could be roughly divided into three groups: (1) *monitoring information sources and student progress*, (2) *doing*, and (3) *practices that include both monitoring and doing*. The practices were carried out at different intensities over different weeks.

A weekly review was also conducted of information on sickness absence and planning systems and on teacher-teacher collaboration. Weekly rhythms of practice and other work-related factors were compared with absences. The comparison was made visually and with the peak detection method. The main result of the project is the methodological process described above, where data from different sources is transformed into a tool for weekly comparisons.



Finally, consideration was given to the use of time series data and peak detection in organizations and future opportunities. There is now a new possibility to develop practices and well-being at work. If in the future the monitoring of weekly rhythms is automatic, the monitoring of the results of development may be continuous as well. Further consideration was given to the source of data and weekly rhythms as a novelty in relation to existing methods and models for studying work and well-being at work.

We were also looking for articles in the research literature that utilized digital trace data, work data. In the References section, 106 articles are grouped into seven groups. The articles found were

- 1) general reflections on big data, trace data and analytics;
- 2) general research methodology such as subjective or objective knowledge, data-driven or model-driven research;
- 3) HR-data exploration;
- 4) suggestions from data scientists for the benefit of work data;
- 5) articles on organizational dynamics, processes and time, and digital trace data that could be applied to "time series of work" information;
- 6) studies of expert working time as an organizational level phenomenon;
- 7) studies using Moodle data, 'virtual learning environment', for 'student analytics', anticipating *student* progress.

At the end, in the last appendix, are the blogs written during the project.

## ALKUSANAT JA KIITOKSET

Työn tekemiseen käytetään tietojärjestelmiä yhä enemmän kaikissa organisaatioissa. Koko toimintaympäristö muuttuu digitaaliseksi, kun myös palvelujen käyttäjät toimivat yhteisillä palvelualustoilla. Organisaatiot kehittävät järjestelmiä, joita käyttämällä toiminnan tavoitteet voitaisiin helpommin saavuttaa. Aikaleimattua ”digital trace dataa” kerääntyy jatkuvasti ja automaattisesti työssä käytettyjen järjestelmien lokitietoihin organisaatioissa.

Meillä on ollut ainutlaatuinen tilaisuus saada yhden ammattikorkeakoulun opettajien Moodlen lokitietoa 63 viikon jaksolta tutkimusaineistoksi ja Työsuojelurahastolta rahoitus ”Työdatan hiljaiset signaalit” eli TYHIS -hankkeelle (2018–2019).

Sitä ennen ehti kuitenkin tapahtua kolme onnekasta sattumaa. Saimme muutama vuosi sitten KEVA:ssa käydessämme vinkin tuosta organisaatiosta. Työpaikka oli jo pitkällä tietovarantojen rakentamisessa. Samoihin aikoihin sain työtoveriksi datatieteilijä Antti Ukkosen ja datalähtöinen ajattelu tuli tutuksi. Aiemmilta vuosilta aineisto/ilmiölähtöinen ajattelu oli jo tuttua, joten datatieteen mahdollisuuksiin oli valmiuksia tarttua.

Antin lisäksi hankkeen suunnitteluun osallistui Työterveyslaitoksella tutkimusinsinööri Andreas Henelius. Toteuttamisvaiheessa datatieteen osaajana työskenteli tutkimusinsinööri Jussi Korpela Antin toimiessa neuvonantajana. Uudet datatieteen haasteet veivät kolme osaajaa mukanaan. Tietojenkäsittelytieteen opiskelija Erkki Sinisalo viimeisteli analyysit ja aloitti uusia analyysikokeiluja tuotepäällikkö Riku Louhimon ohjauksessa. Datatieteen edustajat ovat tehneet datan puhdistuksen, analyysit ja kuvaajat. Kiitän heitä pitkämielisyystään, kun työ- ja organisaatiopsykologian edustajana alati ihmettelin, että mitä sille havaintopisteelle tapahtuu. Tulosten tulkinnat ja kaikki tekstit ovat allekirjoittaneen.

Valtavan suuri kiitos ammattikorkeakoululle ja heidän kehittämisryhmälleen. Erityiskiitos ICT-asiantuntijalle, joka toimitti meille helposti työstettävän aineiston. Tämä ammattikorkeakoulu on edelläkävijän roolissa. Emme tienneet varmasti, mitä voidaan aineistosta saada irti. Kiitos yhteisestä matkasta.

Vielä huomaan taustavoimat Työterveyslaitoksella. Kiitos työtovereille Arja Ala-Laurinaholle ja Vilja Rydmanille tekstin oikoluvusta ja Viljalle kiitos kuvituksista teksteihin. Kiitokset silloiselle johtajalle Jorma Mäkitalolle työdatan ymmärtämisestä ja edistämisestä. Kiitos myös salkkujohtaja Päivi Husmanille kannustuksesta uutta luovaan työhön.

Kiitokset kaikille hankkeeseen osallistuneille ja siihen vaikuttaneille sekä laajemmin kaikille työtä tutkiville ja kehittäville keskustelukumppaneille. Olemme uuden äärellä!

Tiina Kalliomäki-Levanto





## SISÄLLYS

1	Johdanto.....	11
2	Työdata .....	13
2.1	Lokitieto .....	13
2.2	Dataa Moodle-järjestelmästä.....	13
2.3	Datan kontekstina ammattikorkeakoulu ja viikko.....	14
3	Työdata tekstinä .....	16
3.1	Datarivillä henkilökoodi, aikaleima ja käyttötapahtuma .....	16
3.1.1	Moodlen käyttö viikonpäivän ja kellonajan mukaan.....	17
3.1.2	Käyttötapahtumien esiintyvyys.....	17
3.2	Viikoittaiset tekstit .....	19
3.3	Moodlessa käytetty aika .....	20
4	Työ käytäntöinä .....	22
5	LDA työdataa analysoivana Tekoälynä.....	23
5.1	LDA.....	23
5.2	LDA:n, käytäntöjen ja työdatan yhteinen viitekehys.....	23
5.3	LDA:n toteutusvaiheet .....	24
6	Tulososa 1: Käytännöt.....	26
6.1	Käytännöt.....	26
6.2	Käytännöt viikoittain.....	27
7	Dataa muista järjestelmistä .....	29
7.1	HR-järjestelmistä tieto sairauspoissaoloista.....	29
7.2	Suunnittelujärjestelmistä tietoa opetustapahtumista ja yhteistyöstä.....	29
8	Tulososa 2: työn ja sairauspoissaolojen viikkorytmit .....	31
8.1	Sairauspoissaolot.....	31
8.2	Työn ja poissaolojen viikkorytmit .....	32
8.2.1	Opiskelijoiden arviointi ja edistymisen seuranta.....	34



8.2.2	Kurssit ja Moodlen käyttöaika sunnuntaisin.....	35
8.2.3	Yhteydenpito ja käyttöoikeudet.....	36
8.2.4	Moodlen kokonaiskäyttö ja suunniteltu työ.....	37
9	Pohdinta .....	38
9.1	Työdatan ja tekoälyn hyödyntäminen kehittämisessä.....	38
9.1.1	Sairauspoissaolot.....	38
9.1.2	Työn ja poissaolojen rytmi .....	39
9.1.3	Hyödyntäminen jatkossa .....	40
9.2	Työdatan ja tekoälyn hyödyntäminen tutkimuksessa.....	40
9.2.1	Käytännöt ja työn hallinta.....	40
9.2.2	Käytännöt aikasarjatielona ja prosessina.....	41
9.2.3	Työn aikasarjatieloa tarvitsee uutta tutkimuksellista otetta.....	43
9.3	Datan virhelähteet.....	44
	LÄHTEET .....	45
	LIITTEET .....	55



# 1 JOHDANTO

Työn tekemiseen käytetään tietojärjestelmiä yhä enemmän kaikissa organisaatioissa. Koko toimintaympäristö muuttuu digitaaliseksi, kun myös palvelujen käyttäjät toimivat yhteisillä sähköisillä palvelualustoilla. Organisaatiot hankkivat ja kehittävät järjestelmiä, joita käyttämällä toiminnan tavoitteet voitaisiin helpommin saavuttaa. Aikaleimattua ”digital trace dataa” kerääntyy jatkuvasti ja automaattisesti työn tekemisessä käytettyjen järjestelmien lokitietoihin organisaatioissa. Tässä raportissa ”digital trace data” tarkoittaa työn tavoitteiden saavuttamiseen käytettyjen järjestelmien lokitietoihin jääviä jälkiä (trace), jotka muodostavat työstä kertovan toiminnallisen datan, lyhyesti työdatan.

Digitaalisuuden lisäksi myös muista lähteistä kumpuaa jatkuvaa muutosta työn tekemiseen. Työdata ja tekoäly on uusi tiedon tuottamisen menetelmä, jonka avulla muutoksesta saadaan tietoa uudella tavalla, kuten rytmeinä, perustuen konkreettisiin tapahtumiin työssä.

Työdataa kertyy tälläkin viikolla valtavasti työpaikoilla ja odottaa keinoja tulla hyödynnettyksi. Tämän raportin keskeisin tavoite onkin kertoa, miten järjestelmien lokitiedoista on mahdollisuus saada tietoa työstä jatkuvaan toiminnan ja työhyvinvoinnin kehittämiseen. Työdatan analysointi onnistuu, koska tekoälyn kehittäminen ja soveltaminen etenevät myös vauhdilla.

Tässä raportissa kerrotaan Työsuojelurahaston rahoittaman ”Työdatan hiljaiset signaalit” – TYHIS-hankkeen työvaiheista, joilla työelämä tutkimuksen uusi aineistolähde ”digital trace data”, työdata, saatiin käytettävään muotoon. Esittelemme analyysitavan ja periaatteet, joita voidaan soveltaa moniin vastaaviin aineistolähteisiin. Työdataa tarkastellaan organisaation yhteisen tekemisen näkökulmasta.

Työdatassa tapahtumille on aikaleima eli tietoa voidaan tarkastella erilaisissa aikaikkunoissa, joka on uusi mahdollisuus työn tutkimuksen alueella. Valitsimme aikaikkunaksi viikon. Tarkastelemme käytäntöjen toteutumista organisaatiossa viikkorytmeinä.

Hankkeessa meillä oli mahdollisuus saada tietoa myös muista tietolähteistä kuten poissaolo- ja suunnittelujärjestelmistä. Myös näissä järjestelmissä tiedolla oli aikaleima ja voimme tarkastella viikkorytmejä mm. sairauspoissaoloista. Sairauspoissaolotietoa on organisaatioissa kerätty järjestelmiin jo useita vuosia, mutta työdata avasi tien viikoittaiseen tarkasteluun. Näin pystyimme tarkastelemaan työn ja poissaolon viikkorytmejä samanaikaisesti.

Raportissa kerrotaan ensin toiminnallisen työdatan yleisiä piirteitä (luku 2). Hankkeen työdata on yhden ammattikorkeakoulun Moodlen lokitietoihin jäävät 2,8 miljoonaa käyttötapahtumaa 63 viikon jaksolta. Käyttötapahtumia tarkastellaan tekstinä, joka ryhmitellään opettajakohteisiksi viikon jaksoiksi, joista muodostuu 10 210 opeviikkotekstiä. (luku 3).

Työdatan sisältämä tieto työn tekemisestä liitetään työn käytäntöjen viitekehykseen. Käytäntöjä sisältävää tekstiaineistoa analysoivana tekoälynä sovelletaan LDA nimistä aihemallinnusta. LDA analysoi tekstien sisältämien sanojen jakaumia. LDA, käytännöt ja työdata jakavat yhteisen periaatteen havaintojen välisistä suhteista. Rajatussa aineiston osassa, dokumentissa, havainnot (sanat, käytännöt, sanaparit) muodostavat jakaumia.

Tekoäly tunnisti 17 käytäntöä, joista laskettiin viikoittaiset käytännön intensiteetit (luku 6). Viikkotason tarkastelu tehtiin myös sairauspoissaolotiedoista ja suunnittelujärjestelmistä saaduista työn tekemisen tiedoista (luku 7). Käytäntöjen ja muiden työhön liittyvien tekijöiden viikkorytmiä verrattiin poissaolorytmiin. Hankkeen keskeisin tulos on edellä kuvattu menetelmällinen prosessi, jossa eri lähteiden data muokataan viikkotason vertailut mahdollistavaksi välineeksi. Vertailu tehtiin visuaalisesti ja huippujen tunnistusmenetelmällä (luku 8).

Lopuksi pohdittiin aikasarjatiedon ja huippujen tunnistamisen hyödyntämistä organisaatioissa ja tulevaisuuden mahdollisuuksia. Kun tulevaisuudessa rytmien seuranta voi olla automaattista, voi käytäntöjen kehittäminen, esimerkiksi kuormitushuippujen hallinta, ja kehittämisen tulosten seurantakin olla jatkuvaa. Vielä pohdittiin aineistolähdettä ja viikkorytmiä uutuutena suhteessa olemassa oleviin menetelmiin ja malleihin tutkia työtä ja työhyvinvointia (luku 9).

Hankkeessa etsittiin myös tutkimuskirjallisuudesta "digital trace dataa", työdataa, hyödynnäviä artikkeleita. Lähteet -osassa 106 artikkelia on ryhmitelty seitsemään ryhmään. Löydettyt artikkelit olivat

- 1) yleisluontoisia big datan, trace datan ja analytiikan pohdintoja,
- 2) yleistä tutkimuksen metodologiasta kuten subjektiivinen vai objektiivinen tieto, datalähtöinen vai mallilähtöinen tutkimusote,
- 3) HR -datan hyödyntämistä pohtivia,
- 4) ehdotuksia datatieteilijöiden näkökulmasta työdatan hyödyntämiseksi,
- 5) organisaatiodynamiikkaa, prosesseja ja aikaa sekä digital trace dataa käsitteleviä artikkeleita, joita voisi soveltaa "työn aikasarja" tietoon,
- 6) asiantuntijan työaikaa organisaatiotason ilmiönä käsittelevät tutkimukset,
- 7) Moodle-dataa, "virtual learning environment" -dataa, hyödyntävät "learning analytics" teemaiset tutkimukset, joissa *ennakoitiin* opiskelijan edistymistä

## 2 TYÖDATA

### 2.1 Lokitieto

Järjestelmien käytön lokitietoihin jäävä digitaalinen trace data voidaan määritellä teoreettisesti event sequence käsiteparin avulla.

Event eli tapahtuma

- sisältää kaksi osaa: tekemisen (esim. avaa) ja kohteen (esim. tiedosto)
- aikaleiman (milloin tapahtuma oli)
- tunnistetiedon tekijästä (yleensä ihminen).

Event sequence tarkoittaa tapahtumien sarjaa, peräkkäisyyttä tai ketjua sisältäen aikaleiman alussa ja lopussa. Kaksi ajallisesti peräkkäistä tapahtumaa voi jo olla sarja, mutta yhden aikaleiman tapahtuma ei. Lokitiedon ajallisesta tapahtumasta voidaan valita aikaväli, joka on esim. tutkimuksen tavoitteen kannalta merkityksellinen. Voidaan tarkastella esim. tunnin, päivän tai viikon aikaisia tapahtumia.

Event voi olla myös vain tapahtuma eli jonain hetkenä ihminen teki jotain, mutta järjestelmään voi jäädä vain tietona, että jokin tapahtuma tapahtui eikä muuta. Kun lokitiedon tapahtumasta voidaan tunnistaa sekä tekeminen että kohde, on mahdollisuus saada rikasta tietoa järjestelmän käytöstä sekä työstä, johon järjestelmää käytetään. Kun tapahtumat muodostavat ajallisen ketjun, on mahdollisuus tunnistaa erilaisia aikasidonnaisia ilmiöitä kuten rytmejä.

### 2.2 Dataa Moodle-järjestelmästä

Hankkeen keskeisin työhön liittyvä data muodostui yhden ammattikorkeakoulun noin 370 opettajan Moodlen lokitietojen käyttötapahtumista, joka muodosti yhden tietotaulun. Sen lisäksi hankkeen käytössä oli suunnittelu- ja HR-järjestelmistä saatua dataa, jotka muodostivat omat tietotaulunsa. Aineisto kattoi ajanjakson tammikuusta 2017 maaliskuuhun 2018.

Data ei sisällä suoria henkilötunnisteita. Ammattikorkeakoulun asiantuntijat pseudonyymi-soivat datan. Henkilölle annettiin tietotaulukohtainen koodi sekä tietotauluja yhdistävä koodi.

Moodle on virtuaalinen oppimisympäristö, jossa työdata on toiminnallista eikä rakenteellista. Toiminnallisuus tarkoittaa sitä, että digitaalinen jälki jää työn tekemisen myötä automaattisesti lokitietoihin. Henkilö ei erikseen merkitse tekemistään. Tutumpi esimerkki on

sähköpostin käyttö. Käytön useudesta voidaan saada tieto suoraan sähköpostin lokitiedoista. Käyttäjä ei erikseen merkitse mihinkään rakenteeseen, että nyt lähetin sähköpostin, tieto on olemassa automaattisesti lokitiedoissa. Rakenteellinen digitaalinen jälki jää, kun henkilö toiminnan ohessa erikseen merkitsee jotain toiseen järjestelmään tai paikkaan, joka on etukäteen suunniteltu. Puhutaan rakenteellisesta kirjaamisesta. Sillä tarkoitetaan järjestelmää, johon etukäteen on valittu vaiheet ja tiedot, joita odotetaan käyttäjän kertovan työstään ja sen vaiheista.

Toiminnallinen järjestelmä on väline, jolla työtä tehdään. Ilman virtuaalista oppimisympäristöä Moodlea työtä ei saada tehdyksi. Luvuissa 3 ja 6 kerrotaan, millaista toiminnallista tietoa Moodle-datasta tunnistettiin.

## 2.3 Datat kontekstina ammattikorkeakoulu ja viikko

Data on syntynyt ja taltioitunut ammattikorkeakoulun opetustyön kontekstissa. Moodle on virtuaalinen oppimisympäristö, jota opettajat käyttävät opetuksen suunnitteluun, toteutukseen, opiskelijan edistymisen seurantaan, arviointiin ja yhteydenpitoon.

Opetustyö suunnitellaan moduuleittain siten, että yksi moduuli sisältää 8 työviikkoa. Syksyllä on kaksi moduulia ja keväällä kaksi. Kesäkuukausina on myös opetusta. Lukuvuosi rytmittyy opetuksen moduuleihin ja vapaajaksoihin keskimäärin seuraavasti hankkeen tutkimusjaksolla:

Kevään 2017

- a) ensimmäinen moduuli on viikot 2–10, jonka keskellä viikko 9 on vapaajakso,
- b) toinen moduuli on viikot 11–18, jonka keskellä on pääsiäisviikko 15 ja
- c) kesäopetus sijoittuu viikoille 20–34, jonka keskellä on viikot 24–31 vapaajaksona.

Syksyn 2017

- d) ensimmäinen moduuli on viikot 35–43, jonka keskellä viikko 42 on vapaajakso,
- e) toinen moduuli on viikot 44–51 ja
- f) vapaajakso on viikot 52–1.

Kevään 2018

- g) ensimmäinen moduuli on viikot 2–10, jonka keskellä viikko 9 on vapaajakso ja
- h) toisesta moduulista on moduulin aloittava viikko 11.

Työ rytmittyy ammattikorkeakoulussa viikkoina, joten datasta valitaan tarkastelujakson aikaikkunaksi viikko. Viikon valitseminen tarkoittaa sitä, että henkilön tapahtumia tarkastellaan jokaiselta viikolta erikseen. Kun henkilöitä on paljon, "irrationaalisia" henkilöviikko-tietoja



on myös paljon. Tarkastelutasoksi valitaan organisaatio, jolloin tapahtumia ja tietoja tarkastellaan viikolla X kaikkien henkilöiden yhteisen (yhteenlasketun) tiedon perusteella. Yksittäisen henkilön tieto on mukana muodostamassa yhteistä viikoittaista tilannekuvaa työstä.



## 3 TYÖDATA TEKSTINÄ

### 3.1 Datarivillä henkilökoodi, aikaleima ja käyttötapahtuma

Opettajat käyttävät Moodle-järjestelmää opetuksen suunnittelussa ja toteuttamisessa. Opettajien Moodle-järjestelmän käytöstä jäävät tapahtumajäljet lokitietoihin. Lokitiedon sisältö muodostuu kolmesta kokonaisuudesta (taulukko 1): toimijan tunnisteesta (WorkforceIDHashKey), aikaleimasta (ActionTimeStamp) ja käyttötapahtumasta (trgact).

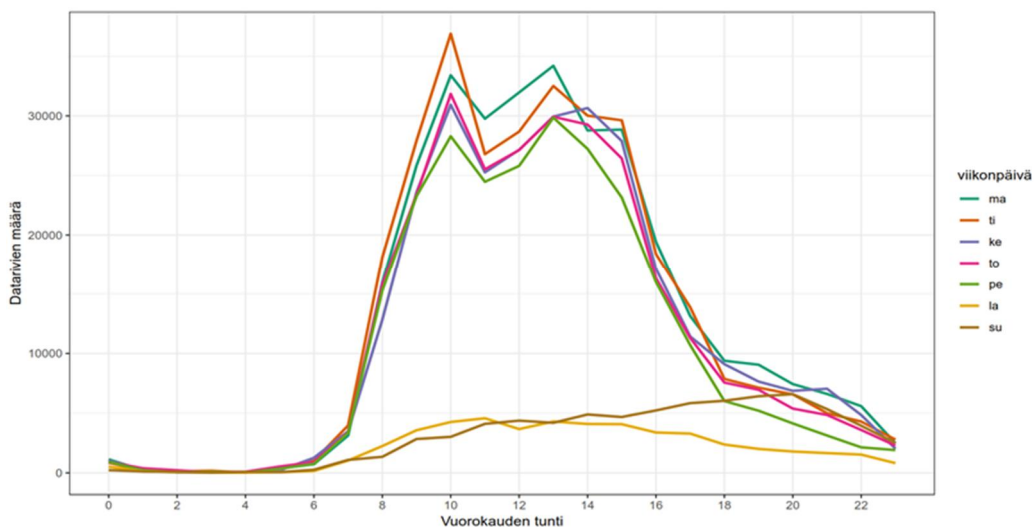
Yksittäinen tapahtuma on määritelty sanaparilla "target-action/trgact" kuten message\_sent, jossa ensimmäinen viittaa tekemisen kohteeseen (message) ja jälkimmäinen tekemiseen (sent). Yksittäisellä kohde-toiminto -käyttötapahtumalla (target-action) on yksiselitteinen aikaleima.

*Taulukko 1. Esimerkkejä Moodlen lokitietojen käyttötapahtumista riveinä.*

WorkforceIDHashKey	ActionTimeStamp	trgact
0e9b734aa25ca8096cb7b56dc0dd8929	2017-01-02 16:06:17	course_module > viewed
a4c42bfd5f5130ddf96e34a036c75e0a	2017-01-02 16:06:42	course > viewed
a4c42bfd5f5130ddf96e34a036c75e0a	2017-01-02 16:07:04	course_module > viewed
a4c42bfd5f5130ddf96e34a036c75e0a	2017-01-02 16:07:07	report > viewed
a4c42bfd5f5130ddf96e34a036c75e0a	2017-01-02 16:07:18	attempt > reviewed
0e9b734aa25ca8096cb7b56dc0dd8929	2017-01-02 16:07:47	course_module > viewed

Moodlessa on rajattu määrä käyttötapahtumia=sanapareja, 185 kappaletta. Henkilö voi esimerkiksi yhtenä viikkona toteuttaa kolmea käyttötapahtumaa usein, osaa vähän ja osaa ei ollenkaan. Osa käyttötapahtumista on siten dominoivassa asemassa suhteessa toisiin joinain viikkoina. Esimerkiksi message\_viewed ja message\_sent yhdistelmä voi esiintyä dominoivana useiden henkilöiden useiden viikkojen tapahtumissa.

Hankkeessa on yhden ammattikorkeakoulun opettajilta 63 viikolta Moodlen lokitiedoista määriteltyä tapahtumadataa tammikuusta 2017 maaliskuuhun 2018. Aineistossa opettajat ovat käyttäneet Moodlen 185 erilaista käyttötapahtumaa yhteensä 2,8 miljoona kertaa eli aineistossa on taulukon 1 mukaisia rivejä 2,8 miljoonaa.



### 3.1.1 Moodlen käyttö viikonpäivän ja kellonajan mukaan

Moodlen käyttö (2,8 miljoonaa riviä) sijoittuivat viikonpäivien ja kellonajan mukaan kuvan 1 osoittamalla tavalla.

*Kuva 1. Moodlen käyttöaktiivisuus datarivien (käyttötapahtumien) määrän avulla kuvattuna viikonpäivien ja kellonajan mukaan*

Moodlen käyttö on aktiivisinta maanantaista perjantaihin kello 8–16 välillä. Sunnuntain iltakäyttö (18–22) on samalla tasolla ma-pe käytön kanssa.

### 3.1.2 Käyttötapahtumien esiintyvyys

Käyttötapahtumista eniten esiintyivät `course_viewed` ja `course_module_viewed` tapahtumat (taulukko 2). Ne tulkittiin olevan rutiininomaisia toimintoja, joita ilman Moodlea ei voi käyttää. (Esim.: opetustunnin alussa opetusmateriaalin jakaminen).

Myöhemmin käytäntöjen jatkoanalyseissa nämä kaksi jätettiin pois. Poisjätettäviksi määriteltiin lisäksi `user_loggedin` ja `user_loggedout`, sillä ko. kirjaukset eivät anna tietoa Moodlen käyttötavasta ja lisäksi kirjauksia puuttui paljon. Jatkoanalyysien kohteeksi rajautui aineisto, jossa on 1,54 miljoonaa riviä. Tämän jälkeen 44 eniten esiintyvää käyttötapahtumaa (taulukko 2) yhteensä kattavat suurimman osan (1,51milj.) aineiston riveistä.



Taulukko 2. Käyttötapahtumien esiintymiskertoja aineistossa (48 eniten esiintyvää 185:sta).

Käyttötapahtuma	Rivien lukumäärä	Käyttötapahtuma	Rivien lukumäärä
course_viewed	812609	question_manually_graded	14400
course_module_viewed	331518	role_assigned	14207
grading_table_viewed	186402	user_enrolment_created	13776
grading_form_viewed	141096	course_bin_item_created	13631
submission_status_viewed	133841	discussion_created	12619
dashboard_viewed	119325	user_profile_viewed	12012
message_viewed	103862	attempt_viewed	10374
user_logged_in	102821	edit_page_viewed	9297
course_module_updated	100729	calendar_event_created	7422
message_sent	85905	course_updated	7209
discussion_viewed	80308	post_created	7144
submission_graded	68243	role_unassigned	6748
user_graded	65741	course_module_incomplete	6561
course_module_created	41456	comment_created	6314
course_section_updated	26717	user_enrolment_deleted	6136
report_viewed	25275	folder_updated	5234
user_list_viewed	25063	group_member_added	5188
assessable_uploaded	23492	page_viewed	5112
attempt_reviewed	22227	discussion_subscription_created	3879
course_module_deleted	20749	grade_deleted	3518
calendar_event_updated	20431	post_updated	3506
grade_report_viewed	16287	course_module_completion_updated	3249
user_logged_out	15769	page_locks_deleted	2910
chapter_viewed	15343	page_updated	2861

Taulukon 2 käyttötapahtumista opiskelijan edistymisen ja tietolähteiden seurantaan (\_viewed) kohdistui 16/44 (36 %) ja tekemiseen (updated, graded, created, deleted ym.) 28/44 (64 %) käyttötapahtumaa.

## 3.2 Viikoittaiset tekstit

Yksittäisten irrallisten 1,54 miljoonan rivin sijaan Moodlen käyttötieto ryhmiteltiin henkilön viikon aikana toteuttamaksi käyttötapahtumajoukoksi eli opeviikko-tiedoksi. WorkforceIDHashKey koodin avulla henkilön viikoittaiset Moodlen käyttötapahtumat voitiin koota yhteen ja muodostaa henkilön viikkotason käyttötapahtumien sanapareista koostuva tekstidokumentti.

Opettajalla on eri viikkoina erilaiset määrät käyttötapahtumia. Esimerkiksi Opettajalla A paljon ja opettajalla B keskimääräisesti käyttötapahtumia. Opettaja A on viikolla 3 (2017) käyttänyt 18 erilaista tapahtumaa yhteensä 2610 kertaa. Opettaja B on viikolla 35 (2017) käyttänyt 19 erilaista tapahtumaa yhteensä 94 kertaa. Esimerkit käyttötapahtumista, niiden määrät ja suhteelliset osuudet eli jakaumat viikossa ovat taulukossa 3.

*Taulukko 3. Esimerkit opettajien A ja B viikoittaisista (2017) käytöntapahtumakertojen lukumääristä ja määrien suhteelliset osuudet eli jakaumat viikossa.*

Käyttötapahtuma target-action	Ope A vko 3 lkm	jakau- ma	Käyttötapahtumat target-action	Ope B vko 35 lkm	jakau- ma
assessable_uploaded	2	0,00	assessable_uploaded	2	0,02
calendar_event_updated	17	0,01	course_module_created	1	0,01
course_mdule_created	1	0,00	dashboard_viewed	13	0,14
course_module_updated	19	0,01	discussion_created	2	0,02
dashboard_viewed	24	0,01	discussion_viewed	1	0,01
discussion_created	2	0,00	grading_table_viewed	10	0,11
grading_form_viewed	892	0,34	message_sent	7	0,07
grading_table_viewed	580	0,22	message_viewed	6	0,06
message_sent	42	0,02	page_created	1	0,01
message_viewed	8	0,00	page_locks_deleted	2	0,02
submission_graded	436	0,17	page_updated	2	0,02
submission_status_viewed	108	0,04	page_viewed	2	0,02
submission_viewed	2	0,00	role_assigned	11	0,12
user_graded	435	0,17	role_unassigned	1	0,01
user_loggedin	15	0,01	submission_status_viewed	14	0,15
user_profile_viewed	1	0,00	user_enrolment_created	11	0,12
dashboard_viewed	24	0,01	user_enrolment_deleted	1	0,01
discussion_created	2	0,00	user_loggedin	5	0,05
			user_profile_viewed	2	0,02
yhteensä	2610	1,00	yhteensä	94	1,00

Opettajalla A viikolla 3 käyttötapahtumat `grading_form_viewed`; `grading_table_viewed`; `submission_graded`; `user_graded` dominoivat kattaen 0,90 suhteellisen osuuden kaikista tapahtumista. Muille tapahtumille jäi yhteensä 0,10 suhteellista osuutta. Opettajalla B viikolla 35 käyttötapahtumat `dashboard_viewed`; `grading_table_viewed`; `role_assigned`; `submission_status_viewed`; `user_enrolment_created` dominoivat kattaen 0,64 suhteellisen osuuden kaikista tapahtumista. Muille tapahtumille jäi yhteensä 0,36 suhteellista osuutta.

Aineiston jatkoanalyysistä rajattiin sellaiset opeviikot pois, joissa oli alle 25 käyttötapahtumaa.

- Analysoitava tekstiaineisto koostuu 440 opettajan 10 210 opeviikkotekstistä, joissa jokaisessa on käyttötapahtumien (sanaparien) jakauma taulukon 5 mukaisesti.
- Moodle-aineiston avulla tutkitaan opetustapahtumien, kuten luennon, ulkopuolelle jäävää opetustyön osuutta.

Opeviikkotekstit muistuttavat aineistona laadullista haastatteluaineistoa.

### 3.3 Moodlessa käytetty aika

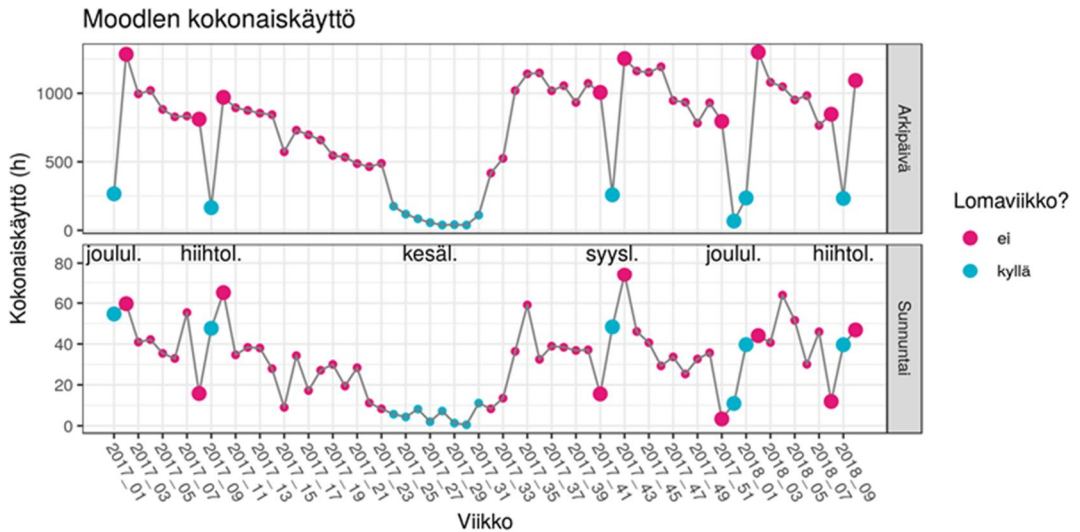
Käyttötapahtumien lisäksi Moodle-datasta arvioitiin Moodlen käyttökertoja. Käyttökerralla tarkoitetaan aikaväliä, jolloin henkilö toimii järjestelmässä aktiivisesti. Käyttökerran kesto minuutteina antaa mahdollisuuden laskea erilaisia käyttöaikoja.

Käyttökerta määriteltiin uuden käyttökerran avulla. Uusi käyttökerta voi alkaa kahdella tavalla. Ensiksi huomioitiin tilanteet, joissa datassa tulee vastaan kirjaus `user_loggedin`. Näitä kirjauksia kuitenkin puuttui ja siksi sovellettiin toista sääntöä, joka perustuu ns. "30 minuutin sääntöön". Jos edellisestä mistä tahansa käyttötapahtumasta on kulunut aikaa enemmän kuin 30 min, niin tällöin seuraava havaittu käyttötapahtuma aloittaa uuden käyttökerran. Jos uudessa käyttökerrassa on vain tuo yksi aloittava käyttötapahtuma ja aikaa kuluu enemmän kuin 30 kunnes tulee seuraava käyttötapahtuma, niin tuo yhden käyttötapahtuman kerta määriteltiin kestoltaan 0 minuutiksi. Vähintään kahden käyttötapahtuman käyttökerrasta saatiin tieto käyttöajasta.

Jokaiselle henkilölle laskettiin jokaiselle viikolle Moodlen käyttökertoista laskettu käyttöaika minuutteina aikavälille maanantai–lauantai ja sunnuntaille erikseen. Moodlen viikon (arki: ma–la ja su erikseen) kokonaiskäyttö saatiin laskemalla henkilöiden käyttöajat yhteen.

Moodlen käyttö arkiviikkoina on lukuvuoden aikana melko tasaista. Syksyllä käyttö on tasan korkealla viikosta toiseen laskien vuodenvaihdetta kohden. Alkukevästä käyttö on syksyn tasolla ja kesää kohden käyttöaika vähenee. Sunnuntaikäyttö vaihtelee (kuva 2).

Sunnuntaikäytöstä havaittiin lomaviikon mukainen rytmi. Tarkastelen seuraavaksi sunnuntaikäyttöä lomaviikkojen ympärillä, sillä näissä esiintyy toistuva rytmi.



Kuva 2: Moodlen kokonaiskäyttö tunteina. Ylemmässä kuvaajassa näkyy kokonaiskäyttö maanantaista lauantaihin ja alemmassa sama sunnuntaille. Lomajaksot on merkitty turkoosilla. Kesän lomajaksot voivat vaihdella opettajakohtaisesti, jolloin loma voi sijoittua toisin, mitä kuvaan on merkitty. Joului-, talvi- ja syyslomat ovat suurimmalle osalle samalla viikolla. Pääsiäinen (2017) on viikolla 15.

Lomaa edeltävän työviikon sunnuntaina Moodlea käytetään vähän, lomaviikon sunnuntaina käyttö lisääntyy ja loman jälkeisen viikon sunnuntaina käyttö edelleen lisääntyy. Viimeksi mainitut sunnuntait ovat usein Moodle käyttömääriltään tarkastelujakson aktiivisimpia ja sitä vastoin lomaa edeltävän viikon sunnuntait kaikkein hiljaisimpia. Näyttäisi siis siltä, että lomaviikon sunnuntaina valmistaudutaan seuraavaan työviikkoon, kun taas lomaviikkoa edeltävä sunnuntai on Moodlen käytön näkökulmasta jo vapaata.

## 4 TYÖ KÄYTÄNTÖINÄ

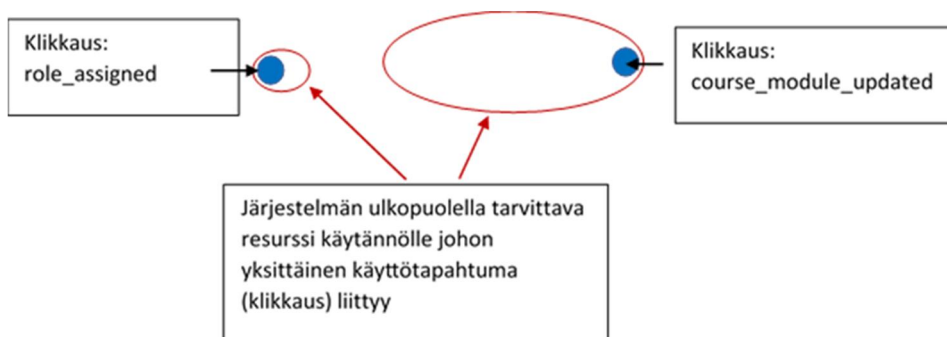
Opetustyö määritellään opetuksen käytännöiksi. Käytännöt ovat ajallis-paikallisia tekojen joukkoja. Moodlen käyttö koostuu ajallis-paikallisista teoista eli käyttötapahtumista jotka samalla ovat opetustyön käytäntöjen osia.

- Käytäntöjen keskinäistä suhdetta voidaan tarkastella mielikuvalla, jossa henkilö "roikottaa" kaikkia työssä tarvittavia käytäntöjä mukanaan, mutta toteuttamishetkellä aktiivisena on vain joitakin.
- Tarkastelun kohteena on systeemin toiminta eikä yksittäinen käytäntö (tai ihminen) irrallaan muista

Yksittäisen tapahtuman käyttämisen määrää ei tässä tarkastella eikä yksittäisen henkilön käyttöä.

Kokonaistoiminnan näkökulmasta käyttötapahtuman klikkaukseen on sitä ennen sitoutunut järjestelmän käytön ulkopuolista resurssia, joka voi olla aikaa, muita järjestelmiä tai henkilöitä. Ulkopuolista resurssia ei tiedetä, mutta siitä tehdään kaksi olettamusta.

- 1) Yksittäiseen käyttötapahtumaan sitoutunut ulkopuolinen resurssi on keskimäärin saman suuruinen eri käyttäjillä. Käyttötapahtuma (klikkaus) voi tarvita vähän tai paljon järjestelmän ulkopuolista resurssia. Käyttäjätunnusten luominen (role\_assigned) voi sitoa vähän edeltävää resurssia (kuvassa vasemmalla) ja kurssien päivitys (course\_module\_updated) paljon (kuvassa oikealla).
- 2) Käyttötapahtuman tarvitsema ulkopuolisen resurssin suhde muihin käyttötapahtumien ulkopuoliseen resurssiin pysyy samana.



Kuva 3. Käyttötapahtumiin voi liittyä vähän tai paljon järjestelmän ulkopuolista resurssia.

## 5 LDA TYÖDATAA ANALYSOIVANA TEKOÄLYNÄ

### 5.1 LDA

Datatieteen parissa on kehitetty tekstin analysointiin LDA-menetelmä, jonka avulla tekstidokumenteista etsitään erilaisia tekstin sisältämiä aiheita/topikkeja/topics. LDA on ohjaimattoman koneoppimisen yksi muoto (Blei, Ng & Jordan, 2003: Latent Dirichlet Allocation).

Tekstissä sanoja on rajallinen määrä. Sanat ovat havaintoyksiköitä, joista LDA etsii sanayhdistelmiä, jotka yhdistelmät toistuvasti esiintyvät tekstidokumenttien joukossa. Analyysi etsii sanajakaumia, jotka ovat keskenään samanlaisia, mutta eroavat muista sanajakumista ja nämä toisistaan erottuvat ryhmittymät nimitetään aiheiksi (topikki/topic).

LDA:n periaatteet sopivat käyttötapahtumien ja käytäntöjen jakaumien analysointiin:

1. tarvitaan dokumentteja, joissa on sanoja
2. oletetaan sanojen muodostavan keskenään jakaumia
3. jakaumat ovat toistuvia aineistossa
4. sanajakaumat ja sanajakumista tunnistettujen aiheiden jakaumat summautuvat ykköseen valittujen dokumenttien rajaamalla alueella.

### 5.2 LDA:n, käytäntöjen ja työdatan yhteinen viitekehys

Mallinnettavan inhimillisen toiminnan kuten käytäntöjen, niihin viittavan datan ja tekoälyn toiminnan taustalla tulisi olla samanlaiset perusolettamukset toiminnan elementtien välisistä suhteista.

LDA:ssa sanat tekstidokumentissa ovat suhteessa toisiinsa eli muodostavat jakaumia ja käytännöt ovat vastaavasti suhteessa toisiinsa ja muodostavat jakaumia. Tieto käytännöistä on käyttötapahtumina Moodlen kohde-toiminto -sanapareina (digital trace data), jotka sanaparit muodostavat jakauman ja analyysin tarvitseman tekstidokumentin (taulukko 4).



Taulukko 4. LDA:n, käytäntöjen ja työdatan yhteinen viitekehys.

Tekoäly Tutkijat ja kehittäjät	Inhimillinen toiminta Tutkijat ja kehittäjät	Työdata Kertyy järjestelmiin
LDA (aihemallinnus)	Käytännöt	Moodlen käyttötapahtumat
Rajallinen määrä sanoja	Rajallinen määrä käytäntöjä	Rajallinen määrä käytäntöön viittavia sanapareja
Sanat muodostavat tekstejä=dokumentteja	Käytännöt muodostavat käytäntökokonaisuuksia rajatussa aikaikkunassa (viikko)	Sanaparit muodostavat tekstejä rajatussa aikaikkunassa (viikko)
Sanat ovat suhteessa toisiinsa ja muodostavat jakauman, joka summautuu ykköseen (yhdessä tekstissä)	Käytännöt ovat suhteessa toisiinsa ja muodostavat jakauman, joka summautuu ykköseen (viikossa)	Sanaparit ovat suhteessa toisiinsa ja muodostavat jakauman, joka summautuu ykköseen (viikon tekstissä)
Tekstejä on paljon	Käytäntöjä toteutetaan paljon	Viikkotekstejä käyttötapahtumista on paljon

### 5.3 LDA:n toteutusvaiheet

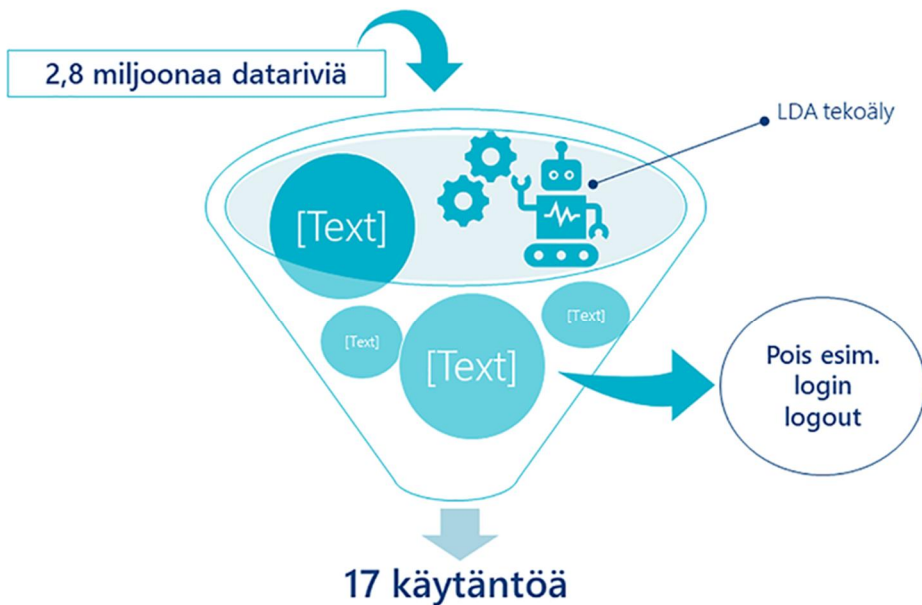
LDA vertaa tekstien (10 210) sanajakaumia toisiinsa ja ehdottaa aineistoon erilaisia aiheiden/topikkien lukumääriä. 17 topikkia saatiin valitsemalla sellainen topikkien lukumäärä, jossa perplexity (ns. osuvuuden mitta) ja topikkien määrä olivat samanaikaisesti mahdollisimman alhaiset. Pienemmän perplexityn omaava LDAn tulos kuvaa paremmin alkupe-  
räisdataa. Koska perplexityllä on tapana laskea topikkien määrää kasvattessa, jouduttiin tyytymään kompromissiin, jossa valittiin topikkien määrä, jonka jälkeen perplexity laski enää vähän (Hoffman, Bach, & Blei, 2010).

LDA vertaa opettajan A viikon X käyttötapahtumien jakaumaa kaikkiin muihin opettajan viikkojakaumiin ja kaikkien muiden opettajien vastaaviin. Vertailulle on valittu sääntönä muodostaa 17 käytäntöä.

LDA ryhmittelee keskenään samanmuotoiset jakaumat omaan ryhmäänsä ja tekee vertailua niin kauan että 17 toisistaan erottuvaa ryhmittelyä löytyy. LDA muodostaa 17 käytäntöjakaumaa, jotka korvaavat opettajan viikon 181 käyttötapahtumien muodostamat jakaumat.

181 käyttötapahtuman sijaan voidaan viikkojen käyttötapahtumien informaatio korvata 17 aiheella/käytännöllä. Opettajan A viikon X 17 käytännöt muodostavat jakauman joka summautuu ykköseen viikolla. Jonain viikkona jokin käytäntö on dominoiva ja muuta käytäntöä on suhteessa vähemmän.

Datasta tunnistetut käytännöt rakentuvat käyttötapahtumien kombinaatioille. Yksittäisen käyttötapahtuman frekvenssi viikolla ei siis ole määrittävä tekijä yksinään, kun LDA on topikkia/aihetta muodostanut. Yksittäisen käyttötapahtuman korkeat frekvenssit datassa ei ole se keskeisin tieto, millä LDA toimintaa mallintaa vaan keskeistä on käyttötapahtumien suhde toisiinsa.



Kuva 4. Tekoäly LDA (ohjaamaton koneoppimisen muoto/aihemallinnus) tiivistää opettajien viikoittaisista 10 210 käyttötapahtumista 17 käytäntöä.

## 6 TULOSOSA 1: KÄYTÄNNÖT

### 6.1 Käytännöt

Käyttötapahtumat ryhmittivät LDA-analyysissä 17 aiheeseen/topikkiin/topic, joita tässä nimitetään käytännöiksi. Jokaisessa käytännössä on teoriassa läsnä kaikki 181 käyttötapahtumaa, mutta vain osa niistä dominoi kerrallaan viikon aikana. Käytännön nimeämistä varten käyttötapahtumista valittiin vähintään 0.15 todennäköisyydellä jakaumassa mukana olevat tapahtumat.

Aihemallinnuksen tuloksena saadut käytännöt (aiheet/topikit/topics) ovat:

- Topic #0: course\_module\_instance\_list\_viewed grade\_report\_viewed
- Topic #1: course\_module\_updated submission\_status\_viewed
- Topic #2: grading\_form\_viewed
- Topic #3: course\_module\_created
- Topic #4: discussion\_viewed
- Topic #5: course\_bin\_item\_created course\_module\_deleted course\_section\_updated
- Topic #6: folder\_updated page\_locks\_deleted page\_updated page\_viewed
- Topic #7: grading\_table\_viewed submission\_graded user\_graded
- Topic #8: role\_unassigned user\_enrolment\_deleted
- Topic #9: attempt\_viewed edit\_page\_viewed
- Topic #10: message\_sent message\_viewed
- Topic #11: chapter\_viewed
- Topic #12: dashboard\_viewed
- Topic #13: grading\_table\_viewed submission\_status\_viewed
- Topic #14: role\_assigned user\_enrolment\_created
- Topic #15: attempt\_reviewed question\_manually\_graded report\_viewed
- Topic #16: comment\_created user\_list\_viewed user\_profile\_viewed

LDA tulostaa "topicit" oman logiikan mukaisessa järjestyksessä.

Käytännöt seurannan ja tekemisen näkökulmasta voitiin ryhmitellä kolmeen ryhmään: 1) vain seurantaa (6/17), 2) vain tekemistä (4/17) ja 3) seurantaa ja tekemistä (6/17). Käytäntöjä, joissa oli seuranta mukana, oli 12/17 (71 %). Opiskelijan edistymisen ja tietolähteiden seuranta on keskeistä suuressa osassa käytäntöjä.



## 6.2 Käytännöt viikoittain

Opettajille lasketaan jokaiselle viikolle omat todennäköisyydet kullekin käytännölle, jolloin saadaan opettajan käytäntöjen todennäköisyysluvut 0–~1 välillä.

Organisaation viikkokohtaiset lukemat yksittäisen käytännön toteuttamisen asteelle saadaan laskemalla yhteen käytännön opettajakohtaiset todennäköisyydet.

Yhteenlaskettua intensiteettiä voidaan havainnollistaa seuraavasti.

- Jos opettaja A toteuttaa 17 käytäntöä kaikkia yhtä paljon viikolla X, on jokaisen käytännön todennäköisyys opelle 1/17 eli 0,06. Jos kaikki opettajat (440) tekevät samoin, saadaan viikolla X intensiivisyyttä kuvaavaksi luvuksi  $440 \times 0,06 = 26,4$ .
- Jos tuolla viikolla 100 opettajaa ei käytä ollenkaan Moodlea ja 340 opettajaa käyttää kaikkia tasaisesti saadaan yhteissummaksi  $340 \times 0,06 = 20,4$ .

Yhteenlaskettu luku kuvaa käytännön yhteistä intensiteettiä organisaatiossa. Käytännöt erosivat toisistaan intensiivisyyden asteen mukaan. Toisia käytäntöjä toteutettiin intensiivisemmin kuin toisia (taulukko 5, liite A).

*Taulukko 5. Käytäntöjen intensiteetin maksimiarvot. Intensiteetti=opettajien yhteenlaskettu todennäköisyys toteuttaa käytäntöä. (\*Topic #12: dashboard\_viewed otettiin käyttöön kesken tutkimusjakson ja siksi sitä ei tulkita).*

Käytäntö (topic/topikki)	Intensiteetin maksimiarvo
<del>Topic #12: dashboard_viewed</del>	<del>56*</del>
Topic #13: grading_table_viewed submission_status_viewed	51
Topic #10: message_sent message_viewed	47
Topic #1: course_module_updated submission_status_viewed	42
Topic #5: course_bin_item_created course_module_deleted course_section_updated	29
Topic #2: grading_form_viewed	23
Topic #4: discussion_viewed	22
Topic #7: grading_table_viewed submission_graded user_graded	17
Topic #3: course_module_created	13
Topic #14: role_assigned user_enrolment_created	11
Topic #16: comment_created user_list_viewed user_profile_viewed	11



---

Topic #9: attempt_viewed edit_page_viewed	9
Topic #15: attempt_reviewed question_manually_graded report_viewed	7
Topic #0: course_module_instance_list_viewed grade_report_viewed	6
Topic #6: folder_updated page_locks_deleted page_updated page_viewed	6
Topic #8: role_unassigned user_enrolment_deleted	4
Topic #11: chapter_viewed	3

---

Seuranta sisältävät käytännöt ovat intensiteetiltään korkeampia (#13, #10, #1) kuin tekemisestä kertovat (#5, #3). Kun Topic #7:n mukaisesti tehtiin arviointia (user graded), niin samalla myös arvioitiin tehtäviä (submission graded) ja seurattiin tietolähteitä (grading table viewed). Voidaan ajatella, että arvioinnin tekeminen on kertaluonteista (lopullista, pakko arvioida). Kertaluonteiset eivät intensiteetiltään nouse arvona korkealle, kun taas seurantaan painottunut käytäntö on jatkuvaa ja toistuvaa, jolloin intensiteettikin on korkea.

Mitä isompi käytännön asteluku, sitä todennäköisemmin käytäntöä toteutetaan. Pieni asteluku kertoo vähäisestä todennäköisyydestä, mutta sen rytmi voi kulkea yhdessä muiden organisaatiorytmien kanssa ja olla kriittinen tekijä työn tavoitteiden saavuttamisessa ja hyvinvoinnin kehittämisessä. Taulukon käytännöistä laaditaan viikkotason kuvaajat, joita esitellään poissaolorytmien kanssa luvussa 8.

## 7 DATAA MUISTA JÄRJESTELMISTÄ

### 7.1 HR-järjestelmistä tieto sairauspoissaoloista

Hankkeen keskeisin datalähde on Moodlen käyttötapahtumat. Sen lisäksi hankkeen käytössä oli suunnittelu- ja HR-järjestelmistä saatua dataa. HR-järjestelmistä saatiin henkilön omaan poissaoloon liittyvää tietoa (taulukko 6). Esimerkiksi ensimmäisellä rivillä on tieto yhdestä poissaolotapahtumasta, joka alkoi 13.11.2017 ja päättyi 15.11.2017. Poissaolon kesto oli 3 päivää ja poissaolon syy sairausloma. Poissaolon tietotaulussa oli 1030 riviä eli poissaolotapahtumaa. Yksi poissaolotapahtuma voi kestää yhden päivän tai monia päiviä, jos on pidemmästä sairauspoissaolosta kyse. Viikolle X laskettiin henkilöiden alkavat lyhyet poissaolot yhteensä.

Taulukko 6. Esimerkkejä poissaolotiedosta.

SPIDHashKey	HashDiff	StartDate	EndDate	Kesto	PoissaolosyyCode	PoissaolosyyTxt
00350e7ab1bc9635 5b49fd04e45e2b22	3dee8437e7305e56 089450446a390729	11.13.2017	11.15.2017	3	20	Sairausloma
01015eebb0d2c8d2 2d18481550f423ba	dba9f5babdb6e029 fa9181f6392b0b28	8.24.2017	8.25.2017	2	20	Sairausloma
01718a69c9c2f2b9a 66b8ef5fdd104d8	cadbb8a02e9889d7 acaf903f501d81b4	9.8.2017	9.14.2017	7	20	Sairausloma

### 7.2 Suunnittelujärjestelmistä tietoa opetustapahtumista ja yhteistyöstä

Suunnittelujärjestelmästä saatiin tieto opettajille suunnitelluista opetustapahtumista (taulukko 7). Esimerkiksi rivillä 1 opettajalla on 11.12.2017 opetustapahtuma, joka alkaa klo 20:00 ja loppuu 22:15 ja kestää 135 minuuttia. Opetustapahtumista arvioitiin kahta asiaa: kesto ja lukumäärää. Viikolle X laskettiin henkilöiden opetustapahtumien lukumäärä yhteensä ja kesto minuutteina yhteensä.



Taulukko 7. Esimerkkejä opetustapahtumista

CNHash-Key	SubjectID-HashKey	StartTime	EndTime	SubjectID	Realization	kes-to	date
16e429920 71601d4e3 733ff4b831 ce12	91e4fcd2 7aa205ef 28f1e08e 0265cc14	11.12.2017 20:00	11.12.2017 22:15	ICT project implementation /ICT Project TK00BN47-3002	TK00BN47-3002	135	11.12.2017
16e429920 71601d4e3 733ff4b831 ce12	91e4fcd2 7aa205ef 28f1e08e 0265cc14	13.12.2017 20:00	13.12.2017 22:15	ICT project implementation /ICT Project TK00BN47-3002	TK00BN47-3002	135	13.12.2017
fb47bb263 9a2df6c3b4 d36ab33a2 74ff	91e4fcd2 7aa205ef 28f1e08e 0265cc14	13.12.2017 20:00	13.12.2017 22:15	ICT project implementation /ICT Project TK00BN47-3002	TK00BN47-3002	135	13.12.2017
d5b2dac9f 10f0ec9f02 a536c5ba2 a29d	91e4fcd2 7aa205ef 28f1e08e 0265cc14	13.12.2017 20:00	13.12.2017 22:15	ICT project implementation /ICT Project TK00BN47-3002	TK00BN47-3002	135	13.12.2017

Samaan tapahtumaan kiinnitetyt opettajat muodostivat yhteistyöverkoston. Yhteistyöverkoston analysoimiseksi käytettiin verkostanalyysiä, jonka avulla henkilölle saadaan arvot keskeisyydestä ja yhteistyön asteesta verkostossa. Yhteistyön aste saadaan yhteistyökumppaneiden määrän perusteella. Liitteessä B on verkostanalyysin periaatteet ja esimerkkikuvia. Viikolle X laskettiin henkilöiden yhteistyön aste yhteensä, mutta keskeisyysarvoa ei käytetty tässä tarkastelussa.

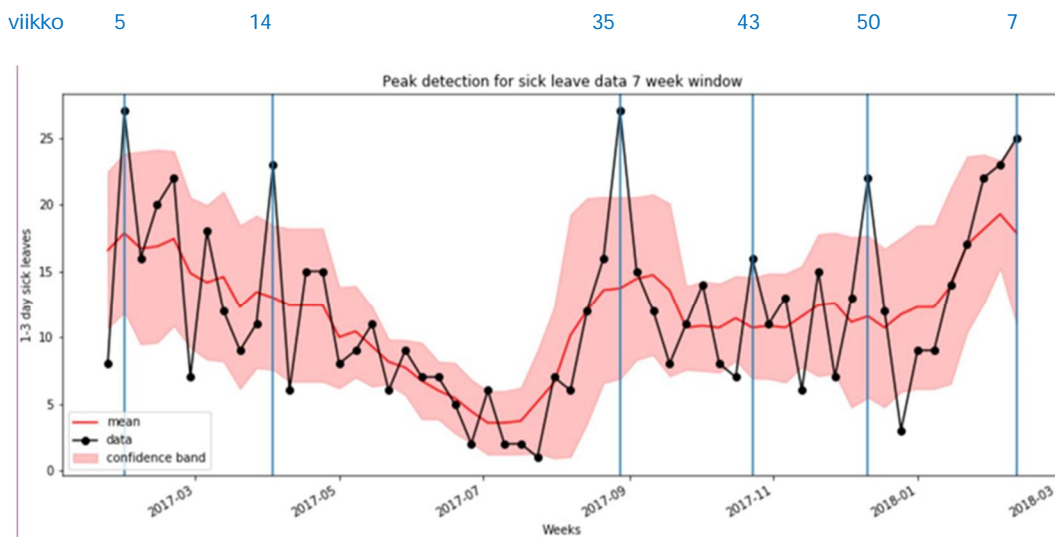
*Moodle ja sairauspoissaolotieto ovat kehkeytyvää tietoa.* Kehkeytyvä, toiminnallinen, tieto syntyy työtä tehdessä. Kehkeytyvää tietoa ei suunnitella etukäteen vaan tieto muodostuu ammattikorkeakoulun opetustyön arjessa. Tieto opetustapahtumista ja niihin liittyvästä yhteistyöstä on suunniteltu valmiiksi eri järjestelmiin. Suunniteltu asia tietona ei synny vasta työtä tehdessä vaan on jo suunnitelmassa.

Moodlen käyttötapahtumat ja poissaolot ovat keskeisin aineistolähde tässä hankkeessa ja niitä voidaan tarkastella data-lähtöisin menetelmin, joissa sallitaan tiedon kehkeytyminen.

## 8 TULOSOSA 2: TYÖN JA SAIRAUSPOISSAOLOJEN VIAKKORYTMIT

### 8.1 Sairauspoissaolot

Aloitetaan viikkorytmin tarkastelu ensin sairauspoissaoloista, joka on ilmiönä tuttu. Aineistosta valittiin 1–3 peräkkäistä päivää olevat henkilön omasta sairaudesta johtuvat poissaolot. Jokaiselle viikolle laskettiin sinä viikkona alkavien poissaolojen määrä poissaolon ensimmäisen päivämäärän perusteella (kuva 5). Tummennettu osa on luottamusväli, johon suurin osa havainnoista sijoittuu. Luottamusväli-kuvaajan luomisessa käytettiin 7 viikon liukuvaa keskiarvoa. Tummennetun osan ulkopuolella olevat huiput ovat poikkeama tuosta joukosta, johon suurin osa havainnoista sijoittuu. Huippujen tunnistus on menetelmä, jolla etsitään keskimääräisistä toistuvista rytmeistä poikkeamia. Kuusi huippuviikkoa on kuvassa merkattu sinisellä pystyviivalla.



Lomaviikko 9 pääs15 24----31 42 52-1

Kuva 5. Alkavat lyhyet (1–3 päivää) kestävät poissaolot viikoittain.

Sairauspoissaoloaineistossa oli kaikkiaan 1030 riviä tietoa sairauspoissaolotapahtumista. 1–3 päivää kestäviä poissaoloja oli 747 eli 75 % kaikista poissaoloriveistä.



## 8.2 Työn ja poissaolojen viikkorytmit

Tässä luvussa esitellään työhön liittyviä viikotason kuvaajia, jotka on laadittu samalla tavalla kuin poissaolokuvaaja edellä kuvassa 5. Työn ja poissaolon kuvaajia tarkastellaan rinnakkain ja verrataan poissaolohuippujen ajankohtia.

Työn käytännöt ryhmiteltiin viiteen ryhmään tutkijan toimesta.

### Opiskelijoiden arviointi ja edistymisen seuranta

Topic #7: grading\_table\_viewed submission\_graded user\_graded

Topic #13: grading\_table\_viewed submission\_status\_viewed

Topic #2: grading\_form\_viewed

Topic #9: attempt\_viewed edit\_page\_viewed

Topic #15: attempt\_reviewed question\_manually\_graded report\_viewed

### Kurssit

Topic #1: course\_module\_updated submission\_status\_viewed

Topic #5: course\_bin\_item\_created course\_module\_deleted course\_section\_updated

Topic #3: course\_module\_created

### Yhteydenpito

Topic #10: message\_sent message\_viewed

Topic #4: discussion\_viewed

### Käyttöoikeudet ja ilmoittautuminen

Topic #14: role\_assigned user\_enrolment\_created

Topic #8: role\_unassigned user\_enrolment\_deleted

### Muut

Topic #12: dashboard\_viewed

Topic #16: comment\_created user\_list\_viewed user\_profile\_viewed

Topic #0: course\_module\_instance\_list\_viewed grade\_report\_viewed

Topic #6: folder\_updated page\_locks\_deleted page\_updated page\_viewed

Topic #11: chapter\_viewed

Käytäntöjen kuvaajat laadittiin kaikista muista, paitsi ryhmään "muut" kuuluvista käytännöistä. Viikkokuvaajat laadittiin myös yhteistyön asteesta, opetustapahtumien määrästä ja kestosta. Vertailussa oli lisäksi mukana edellä kuvassa 2 esitetty Moodlen viikoittainen käyttöaika sunnuntaina sekä Moodlen käyttötapahtumien kerrat kaikkiaan yhteensä kaikilta opettajilta. Työn kuvaajia tulkitaan omien huippujen osalta samalla tavalla kuin poissaoloja. Kun käytännön viikon lukuarvo on tummennetun alueen ulkopuolella, on kyse huipusta, joka erottuu keskimääräisestä tapahtumisesta (tummennetulla alueella).

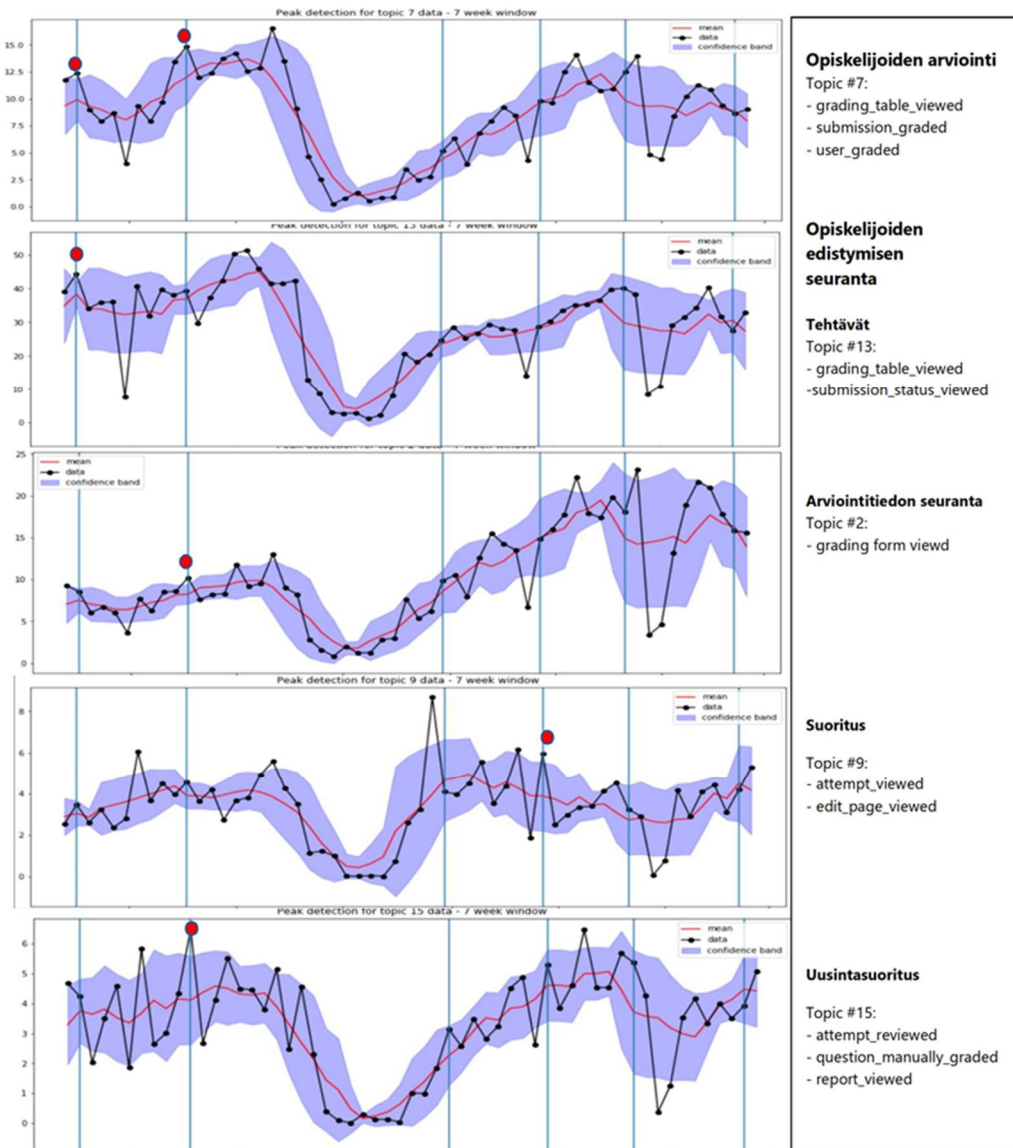


Kuvaajiin on piirretty sininen viiva poissaolon huippuviikkojen kohdalle, jotta voidaan pohdita, mitä poissaolohuippujen ympärillä ammattikorkeakoulun käytännössä tapahtuu. Punainen ympyrä on niillä työviikoilla, joiden huiput osuvat samaan poissaolon huippuviikkojen kanssa. Lomaviikot ovat 2017: 9, 2–31, 42, 52–1, 2018:9 ja pääsiäinen viikko 15.



## 8.2.1 Opiskelijoiden arviointi ja edistymisen seuranta

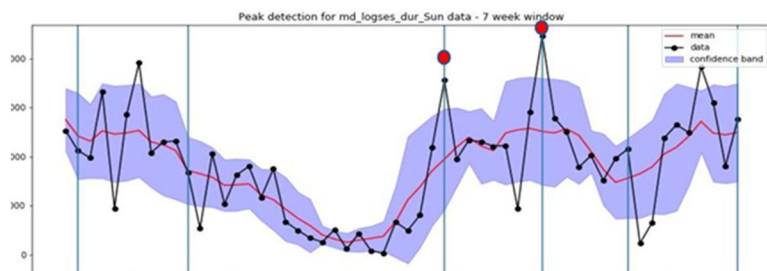
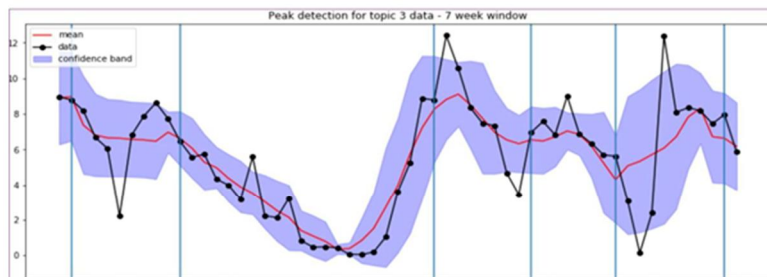
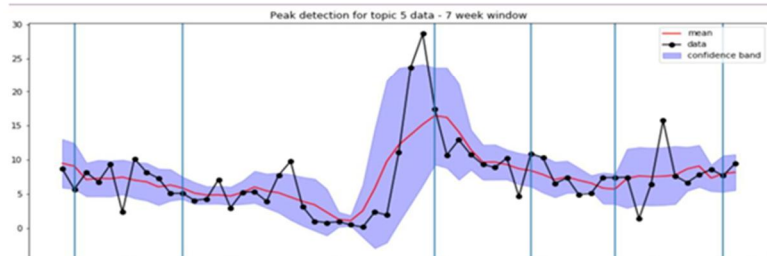
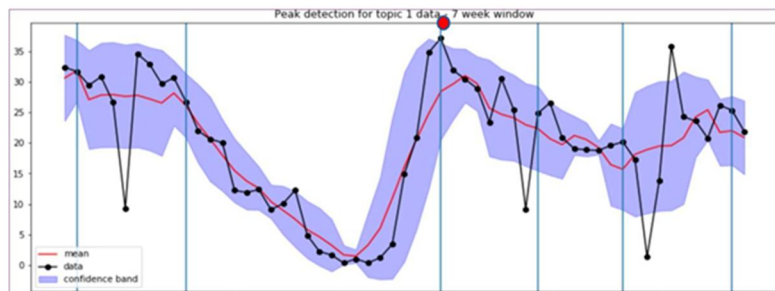
viikko 5 14 35 43 50 7



Arvioinnin huiput osuvat yksiin alkukevään poissaolohuippujen kanssa.

## 8.2.2 Kurssit ja Moodlen käyttö sunnuntaisin

viikko 5 14 35 43 50 7



### Kurssit

#### Kurssien päivitys

Topic #1:  
- course\_module\_updated  
- submission\_status\_viewed

#### Kurssien poisto

Topic #5:  
- course\_bin\_item\_created  
- course\_module\_deleted  
- course\_section\_updated

#### Kurssien luominen

Topic #3:  
- course\_module\_created

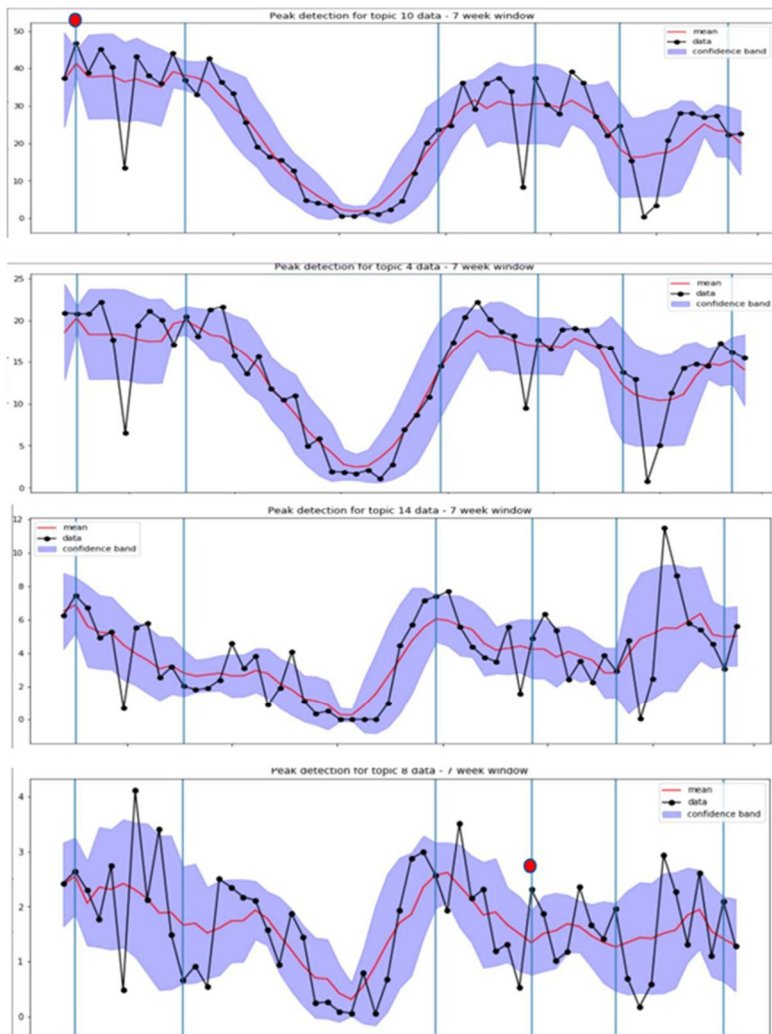
### Moodlen käyttö sunnuntaisin

Erikseen laskettujen Moodlen käyttökertojen ajallinen kesto yhteensä minuutteja sunnuntaisin

Kurssien päivitys osuu yksiin syksyn poissaolon huippuviikon 35 kanssa. Syksyn poissaoloihin osuu myös Moodlen käyttö sunnuntaisin.

## 8.2.3 Yhteydenpito ja käyttöoikeudet

viikko 5 14 35 43 50 7



### Yhteydenpito

#### Yhteyden pito viestein

Topic #10:

- message\_sent
- message\_viewed

#### Keskustelun seuranta

Topic #4:

- discussion\_viewed

### Käyttöoikeudet

#### Käyttöoikeudet ja ilmoittautuminen

Topic #14:

- role\_assigned
- user\_enrolment\_created

#### Käyttöoikeuksien ja ilmoittautumisen poisto

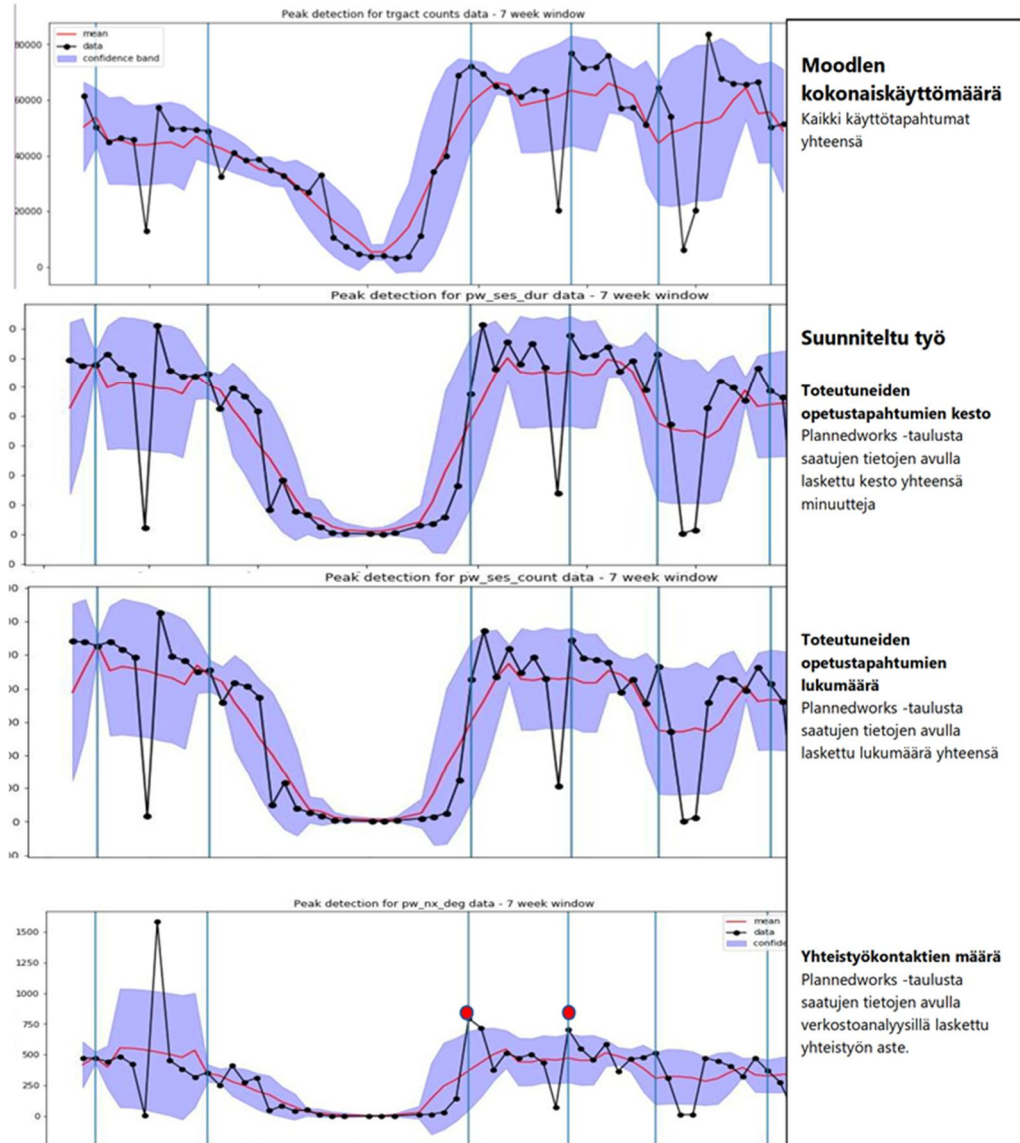
Topic #8:

- role\_unassigned
- user\_enrolment\_deleted

Yhteydenpito ja keskustelun seuranta nousevat syksyn myötä korkealle tasolle jatkuen alkukeväälle ja laskien kesää kohden.

## 8.2.4 Moodlen kokonaiskäyttö ja suunniteltu työ

viikko 5 14 35 43 50 7



Moodlen kokonaiskäyttö ja suunniteltu työ kulkevat intensiteetiltään samassa tahdissa. Yhteistyön aste puolestaan osuu yksiin syksyn poissaolohuippujen kanssa.

## 9 POHDINTA

### 9.1 Työdatan ja tekoälyn hyödyntäminen kehittämisessä

Työdata on uusi aineistolähde, joka kerääntyy jatkuvasti ja automaattisesti järjestelmiin ja sähköisiin palvelualustoihin. On mahdollisuus saada tietoa työstä jatkuvasti ajankohdasta toiseen rytmeinä tai prosesseina. On mahdollisuus saada tietoa, joka on objektiivista ja konkreettista tietoa työn tekemisestä. Tähän asti dataa on pääsääntöisesti kerätty kyselyillä ja saatu tietoa koetuista työn piirteistä poikkileikkauksenomaisesti toistoina esimerkiksi kahden vuoden välein.

Työdata avaa mahdollisuuden käyttää tekoälyä ensiksikin työn konkreettisen tekemisen mallintamiseen ja toiseksi työn mallintamiseen prosessina tai rytmeinä. Kun tieto on valmiiksi työn tekemisestä kertovaa, on sen avulla helpompi suunnitella kehittämistoimia. Jos on mahdollisuus jatkuvaan työdatan ja tekoälyn hyödyntämiseen, kehittämistoimien vaikeavuutta on mahdollisuus seurata tuon jatkuvasti täydentyvän tilannekuvan avulla. Esimerkiksi, jos halutaan muuttaa viikkoja, joilla jokin käytäntö viittaa kuormitushuippuun, voidaan muuttaa käytännön organisointia niin, että huippu madaltuu.

Kun työdata tekee mahdolliseksi tarkastella työtä ja sen käytäntöjä viikon rytmeinä, avautuu samalla mahdollisuus tarkastella myös muita työhön liittyviä tekijöitä kuten poissaoloja ja suunniteltua työtä viikon rytmeinä. Organisaation henkilöstöryhmät voivat yhdessä tulkita esimerkiksi työn intensiteetin huippuja. Ovatko käytäntöjen huiput tuottamassa ylikuormitusta? Millaiset intensiteettiyhdistelmät ovat hankalinta hallita?

#### 9.1.1 Sairauspoissaolot

Sairauspoissaolotietoja on työpaikoilla koottu kauan, eivätkä ne sinällään ole uutta tietoa. Yleensä poissaolotietoja tarkastellaan henkilötyövuosina vuodessa. Tässä tutkimuksessa poissaoloja tarkasteltiin kehittyvänä ilmiönä. Poissaolot tapahtuvat erilaisten tekijöiden myötävaikutuksesta ja muodostavat oman prosessin, jota on mahdollisuus tarkastella myös viikkotasolla. Lyhyet poissaolot ovat itsessään ilmiö, joka tapahtuu kaikissa organisaatioissa. Lyhyiden poissaolojen oletettiin olevan reaktio työhön. Toisaalta henkilö, jolla on eri syistä vähemmän resursseja käytössään (omia tai työpaikan) juuri kuormittavissa ajankohdissa, on myös haavoittuvimmassa tilanteessa, ja hän voi ylikuormittua.

Amerikkalaisessa tutkimuksessa oli käytettävissä iso poissaoloja koskeva data-aineisto. Tutkimuksessa tarkasteltiin viikko-päivä rytmejä sekä tunnistettiin poissaolon huippuviikoja vastaavasti kuin tässä hankkeessa tehtiin (Spears et al., 2013).

Poissaolohuippujen kohdalla voidaan pohtia, mitä niiden ympärillä ammattikorkeakoulun käytännöissä tapahtuu. Tapahtumat voivat olla muutakin kuin työdatasta saatua tietoa. Mitä muuta lisääntyvien poissaolojen lisäksi lisääntyy ja vastaavasti vähenee? Millaisia yhteisiä tulkintoja voidaan tehdä? Jos tarkastellaan olemassa olevia työhyvinvointimalleja kuten työn hallinnan, työn tuunaamisen, palautumisen tai osallistavan esimiestyön malleja, mitä viikkotason tarkastelu niiden näkökulmasta tarkoittaa?

### 9.1.2 Työn ja poissaolojen rytmi

Kun tarkastellaan kokonaiskuormitusta eli suunniteltuja opetustapahtumia ja Moodlen kokonaiskäyttöä käyttötapahtumien lukumääränä vuoden aikana, alkusyksy ja alkukevät ovat kuormitettuja ja kevättä kohden kuormitustaso alenee. Suunnitellut opetustapahtumat ja Moodlen kokonaiskäyttö kulkevat samassa tahdissa. Kuormitus näyttää kulkevan samassa rytmissä poissaolorytmin kanssa. Kun kuormitustaso laskee, laskee myös poissaolot.

Käytännöistä eniten samassa rytmissä kokonaiskuormituksen ja poissaolojen yleisen tason kanssa kulki *yhteydenpito viestein ja keskustelun seuranta*.

Kursseihin ja arviointiin liittyvillä käytännöillä oli omanlaisensa rytmit, joista osa osui poissaolon huippuviikkoihin. Kursseihin liittyvän tekemisen intensiteetti nousi jyrkästi syksyllä, kuten myös syksyn suunnitellut opetustapahtumat. *Kurssien päivituksen* huippuviikko osui samaan poissaolon huippuviikon (vko 35) kanssa.

Syksyn kokonaiskuormituksesta kertoo myös tieto *Moodlen sunnuntaikäytöstä*. Moodlen sunnuntaihuippuviikkoja oli 4, joista 3 syksyllä. Näistä 2 osui alkusyksyn poissaolon huippuviikkoihin.

Kursseihin liittyvän tekemisen huippuviikot osuivat keskenään peräkkäin alkusyksyyn. Ensin poistettiin kursseja viikolla 34, sitten päivitettiin ja viikolla 36 luotiin uusia.

*Opiskelijoiden edistymisen seuranta ja arviointi* -käytäntöjen intensiteettihiiput osuivat samaan aikaan poissaolojen huippujen kanssa alkukevästä. Kesää kohden arvioinnin intensiteetti edelleen hieman nousi, mutta tuo nousu ei kehkeyttänyt poissaoloa, vaan tuona ajankohtana poissaolot vähenivät kesää kohden. Laskuun oli yhteydessä se, että samana ajankohtana loppukevästä sekä suunniteltu työ että muut käytännöt intensiteetiltään vähenivät, jolloin kokonaiskuormituksella oli vähenevä suunta.

Työn ja poissaolon huippuviikkoja voidaan tarkastella myös riskinäkökulmasta. Ovatko jotkin ajankohdat riskiajankohtia työn sujumisen ja hyvinvoinnin näkökulmasta? Miten eri työuran vaiheissa huippuviikot vaikuttavat?

Organisaation henkilöstöryhmät voivat yhteistyössä tehdä tarkemmat tulkinnot viikoittaisista tilannekuivist. Mihin tulisi kiinnittää huomiota toiminnan suunnittelun ja johtamisen näkökulmasta? Entä mitä HR ja työsuojelun näkökulmasta tulisi vielä tarkemmin tietää?





Mikä opettajien ja oppilaiden näkökulmasta on tärkeää? Miten ICT-järjestelmiä kehittävät ovat mukana tiedon tuottamisessa? Mitä yhdessä kehitetään? Onko epävarmuutta aiheuttavia asioita? Vaarantuuko jokin tällä hetkellä hyvin toimiva tai muutoin tärkeä asia, jos ryhdytään tekemään muutosta? Millainen muutos on mahdollista? Miten palautuminen onnistuu intensiivisissä jaksoissa?

### 9.1.3 Hyödyntäminen jatkossa

Tässä hankkeessa tutkittiin tarkkaan yhtä aineistoa ja saatiin algoritmeja, joita voidaan jatkossa soveltaa melko helposti vastaaviin ammattikorkeakoulun aineistoihin. Moodlea vastaavaa työdataa muilta toimialoilta ja niiden järjestelmistä on myös helpompi lähteä työstämään tässä raportissa esitelyjen vaiheiden kautta. Toimialakohtaisesti voisi mallintaa yksittäisiä organisaatioita ja sitten laajentaa hyödyntäminen useampiin alan organisaatioihin.

Yhteinen, käytännön työn tekemiseen liittyvä tilannekuva viikoittaisena rytminä on saatavilla kaikille henkilöstöryhmille organisaatiossa. Kun tieto on jatkuvasti kertyvää, voidaan tulevaisuudessa kehittää algoritmeja niin, että ne voisivat hyödyntää organisaatiota jatkuvassa oman tilanteen seurannassa.

Tulevaisuudessa tilannekuva voi olla jatkuvasti päivittyvä. Näin ollen on mahdollista nopeasti tunnistaa kehitettävät kohdat, kokeilla toisella tavalla tekemistä ja saada nopeasti palaute ja jatkaa kehittämistä.

Tässä hankkeessa työhyvinvointia on tarkasteltu sairauspoissaolojen avulla. Toiminnan sujumista ja hyvinvointia voi tarkastella laajemmasta näkökulmasta. Työdatan lisäksi saadaan palvelun käyttäjistä tietoa. Ammattikorkeakoulussa laajempi näkökulma tarkoittaisi opiskelijoiden vastaavien viikkotapahtumien mukaan ottamista tutkimusmalliin.

Kun organisaatiot omistavat työdatan, joka on valmiiksi kerääntynyt lokitietoihin, ei tarvita erillistä tiedon keruuta, vaan työdatasta saadaan algoritmien avulla toiminnan suunnitteluun ja johtamiseen tietoa. Useamman organisaation työdatasta saadaan alakohtaisia trendejä, joita voidaan julkaista Työterveyslaitoksen työelämä-tieto -palvelussa.

## 9.2 Työdatan ja tekoälyn hyödyntäminen tutkimuksessa

### 9.2.1 Käytännöt ja työn hallinta

Käytännöt ovat organisaatiossa ajallis-paikallisia tekojen joukkoja, jotka ovat muodostuneet toiminnan tavoitteen saavuttamista varten (Feldman & Orlikowski 2011). Yksittäinen Moodlen käyttötapaus on ajallis-paikallinen teko ja osa käytäntöä.

Yksilötasolla käyttötapahtuma tulkitaan toiminnan teorian avulla, jossa on kolme tasoa. Ihmisen toiminta on 1) tavoitteeseen suuntautuvaa käyttäytymistä, jota toteutetaan 2) tekojen ja 3) operaatioiden avulla. (Frese & Zaph 1994). Käyttötapahtumat ovat ihmisen toiminnan säätelyn näkökulmasta operaatioita. Yksittäinen käytäntö, esimerkiksi "message\_sent message\_viewed" on ihmisen toiminnan säätelyn kannalta teko. Kaikkia opetuksen operaatioita ja tekoja ohjaa toiminnan kokonaistavoite, joka on oppilaiden edistyminen ja oppiminen.

Tässä hankkeessa tunnistettujen opettajien viikkokäytäntöjen keskeisin piirre oli opiskelijan edistymisen tai tietolähteiden seuranta. Seuranta sisältyi ainakin osana 71 % käytännöistä. Eniten käytetyistä *yksittäisistä* käyttötapahtumista seurantaan kohdistuvia oli kuitenkin vain 36 %. Työviikkoja erottelevaksi tekijäksi nousikin käytännöt, joissa seuranta oli eri muodoissa mukana.

Työn vaatimukset–työn hallinta -malliin (Karasek 1979) on mahdollisuus saada paljon uusia elementtejä työdatasta.

- Mahdollisuus toiminnan seurantaan antaa keinon ennakointiin, joka puolestaan voisi olla yksi uusi työn hallinnan teoreettinen elementti.
- Seuranta ja ennakointi Moodlen avulla ja tuon seurannan tunnistaminen tietoiseksi työn hallinnan välineeksi voisi edistää kuormituksen hallintaa.
- Moodlen kehittäminen siihen suuntaan, että seuranta onnistuu helposti.
- Jos työn vaatimukset ovat käytäntöjen huippuja, on mahdollisuus vaikuttaa myös vaatimukseen suoraan sen lisäksi, että tunnistetaan ja kehitetään hallintaa. Karasekin (1979) perusajatus oli, että työn vaatimusta/kuormitusta tarvitaan ja sillä saadaan työn tulos aikaiseksi. Nyt on mahdollisuus tunnistaa ylikuormitusta suoraan työn tekemisestä, mitä aiemmin ei ole ollut mahdollista.

### 9.2.2 Käytännöt aikasarjatietona ja prosessina

Oletamme, että käytäntöjä on mahdollisuus tunnistaa TYHIS-hankkeessa kehitetyin menetelmin monista virtuaalisista työn tekemisen järjestelmistä. Tarvitaan taulukon 1 mukaiset tiedot eli henkilö koodina, aikaleima ja käyttötapahtuma, jossa on tekeminen ja sen kohde. Tuloksena saadaan käytäntöjä, jotka ovat konkreettisia tapahtumia työn tekemisessä. Käytännöille on olemassa teoreettinen tausta tutkimuskirjallisuudessa, mutta käytännöt prosessina ja "aikasarjana" organisaatiotason ilmiönä ovat uutuus, jolle ei vielä löydy hyväksytyjä analyysitapoja, saati teoreettisia malleja.

Prosessi tarkoittaa tässä hankkeessa

- aikaleimatusta työdatasta tunnistettuja käytäntöjä ja
- käytäntöjen tapahtumista ajankohdasta toiseen – tässä hankkeessa viikosta toiseen.

Vastaavia prosesseja on meneillään ja niitä voidaan analysoida monissa ammattikorkeakouluissa ja tietointensiivisissä työorganisaatioissa, joissa käytetään järjestelmiä.

Tapahtumat prosessimaisten ilmiöiden tutkimisessa voidaan jakaa neljään ryhmään:

- 1) Tapahtumat prosessissa vain sattuvat peräkkäin.
- 2) Tapahtumat muodostavat rytmin – on olemassa episodinen rytmi, joka toistuu.
- 3) Tapahtumat kertovat muutoksesta, jossa oletetaan erityisen eventin/tapahtuman olemassaolo. Event voi olla esimerkiksi iso organisaatiomuutos. Muutoksella on vaikutusta prosessiin. Muutoksen vaikutusta seurataan ennalta määritettyjen muuttujien avulla.
- 4) Tapahtumat kertovat jonkin tekijän kehittymisestä ajan kuluessa. Tutkimuksessa on voitu laatia hypoteesi aiemman tiedon perusteella, jonka mukaan ilmiö kehittyisi tietyllä tavalla, ja havaittavien tapahtumien kautta saadaan tieto kehittämisestä. (Kouamé & Langley, 2018)

Tämä tutkimus on esimerkki kohdasta 2. Emme oletta, että jokin iso muutos on tapahtunut, jonka vaikutusta seurataan (3) tai että on jokin ilmiö, jonka kehityskaarta tulisi seurata (4). Emme myöskään oletta, että viikoittaiset tapahtumat, käytännöt, olisivat sattumanvaraisia (1).

Organisaatioiden tutkijat määrittelevät organisaation tapahtumat dynaamisiksi ajassa muuttuviksi eikä staattisiksi ilmiöiksi. Ilmiöitä tulisi tutkia tutkimusasetelmassa, jossa tutkimushavainnot kerätään useasta ajankohdasta. Nyt ehdotamme työdatasta tunnistettua aikasarjatietoa uudeksi mahdollisuudeksi tutkimuksiin.

Tutkijat eivät vielä ole esitelleet artikkeleita, joissa olisi raportoitu, miten ”digital trace data” kertyy työn tekemisen lomassa ja miten työdataa olisi analysoitu tai millaisia viitekehyksiä on käytetty prosessien näkökulmasta. Tosin Feldman, Pentland, D’Adderio, & Lazaric (2016) toteavat, että ajan oloon tulee käyttöön ”digital trace data” ja ”computational methods” uutena keinona tutkimukselle.

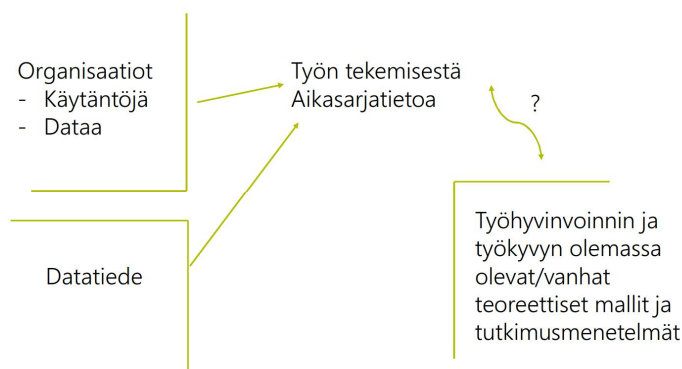
### 9.2.3 Työn aikasarjatieto tarvitsee uutta tutkimuksellista otetta

Nyt on saatavilla työn tekemisestä aikasarjatietoa. Mitä teoreettisia malleja on olemassa, johon työn aikasarjatieto sopisi? Tulisiko kehittää erillistä analytiikkaa? Edellä totesimme, että LDA, käytännöt ja työdata jakavat yhteisen periaatteen havaintojen välisistä suhteista (havainnot muodostavat jakaumia). Vastaavanlainen yhteinen periaate tai viitekehys tarvitaan työn aikasarjalle (taulukko 8).

Taulukko 8. Mikä voisi olla aikasarja-analyysin, organisaatiomallien ja työdatan yhteinen viitekehys?

Tekoäly Tutkijat ja kehittäjät	Inhimillinen toiminta Tutkijat ja kehittäjät	Työdata Kertyy järjestelmiin
Aikasarja-analytiikka	Organisaatiomalit	Data
??	työn, terveyden ja organisaatioilmiöiden aikasarjat <ul style="list-style-type: none"> <li>- peräkkäisyys</li> <li>- episodi</li> <li>- muutos</li> <li>- kehitys</li> </ul>	inhimilliset aikasarjat aikaleimatussa lokitiedossa - paljon vaihtoehtoja

Kaikkiaan on kyse isommasta tutkimusotteen muutoksesta. Tulevaisuuden haasteena on se, että olemassa olevat työhyvinvoinnin teoreettiset mallit ja tiedonhankintamenetelmät perustuvat staattiseen pistemäisesti kerättyyn tietoon, kun muutoksessa tarvitaan dynaamista tietoa, jota työn tekemisen aikasarjatieto on. Miten edetä uusien mallien rakentamisen kanssa?



Kuva 6. Tulevaisuuden haasteet: Miten edetä uusien aikasarjatietoon perustuvien mallien kanssa?

### 9.3 Datan virhelähteet

Datan luotettavuutta on tarkasteltu seuraavista virhettä aiheuttavista näkökulmista. Virheiden alkuperä voi olla tekninen. Kun järjestelmää kehitetään voi syntyä aiemmasta poikkeavaa trace:a, kuten uuden käyttötapahtuman luominen. Hankkeen aineistossa löytyi yksi sellainen käyttötapahtuma, joka otettiin käyttöön tutkimusjaksolla. Käyttötapahtuma huomioitiin, mutta sitä ei tehty tulkintaa. Lokitietojen hallinnoinnista voi tulla virhettä kuten duplikaatteja. Hankkeessa tunnistettiin duplikaatit ja ne poistettiin. Järjestelmän käyttäjistä voi aiheutua teknisiä tai määrittelyyn liittyviä virhelähteitä. Hankkeen aineistossa loggedin ja loggedout jälkiä tuntui puuttuvan, joten niitä ei käytetty ajallisen käyttökerran määrittelyssä (Østerlund, Crowston, & Jackson, 2019).

Datan LDA:n analyysimalli pitää sisällään virheen hallintaa. Jakaumien tarkastelussa yksittäisten käytäntötapahtumien ylliedustus tai puuttuminen joissakin yksittäisissä viikkodokumenteissa hukkuu 10 210 jakauman vertailussa. Jakauman samanmuotoisuutta on tekemässä useat jakaumat, jolloin yksittäisen jakauman mahdollinen virheellisyys ei pääse vaikuttamaan paljoa. Yhden jakauman merkitys on 1/10210.

Aineisto rajattiin yli 24 käyttötapahtumaa sisältäviin teksteihin. Analyysimalli oli hankkeessa uusi. Ei ollut varmaa tietoa siitä, miten LDA toimii, jos mukana on tekstejä, joissa on kovin harvoja sanoja (käyttötapahtumia).

## LÄHTEET

Hankkeessa etsittiin tutkimuskirjallisuudesta ”digital trace dataa”, työdataa, hyödyntäviä artikkeleita. 106 artikkelia on ryhmitelty seitsemään ryhmään.

- 1) yleisluontoisia big datan, trace datan ja analytiikan pohdintoja,
- 2) yleistä tutkimuksen metodologiasta kuten subjektiivinen vai objektiivinen tieto, datalähtöinen vai mallilähtöinen tutkimusote,
- 3) HR -datan hyödyntämistä pohtivia,
- 4) ehdotuksia datatieteilijöiden näkökulmasta työdatan hyödyntämisestä,
- 5) organisaatiodynamiikkaa, prosesseja ja aikaa sekä digital trace dataa käsitteleviä artikkeleita, joita voisi soveltaa ”työn aikasarja” tietoon,
- 6) asiantuntijan työaikaa organisaatiotason ilmiönä käsittelevät tutkimukset,
- 7) Moodle-dataa, ”virtual learning environment” -dataa, hyödyntävät ”learning analytics” teemaiset tutkimukset, joissa ennakoitiin *opiskelijan* edistymistä.

- 1) Yleisluontoisia big datan, tracedatan ja analytiikan pohdintoja

Berente, N., Seidel, S., & Safadi, H. (2018). Research Commentary—Data-Driven Computationally Intensive Theory Development. *Information Systems Research*. doi:10.1287/isre.2018.0774

George, G., Haas, M. R., & Pentland, A. (2014). Big Data and Management. *Academy of Management Journal*, 57(2), 321-326. doi:10.5465/amj.2014.4002

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big Data and Data Science Methods for Management Research. *Academy of Management Journal*, 59(5), 1493-1507. doi:10.5465/amj.2016.4005

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 2053951714528481. doi:10.1177/2053951714528481

Jones, M. (2019). What we talk about when we talk about (big) data. *The Journal of Strategic Information Systems*, 28(1), 3-16. doi:https://doi.org/10.1016/j.jsis.2018.10.005

Østerlund, C., Crowston, K., & Jackson, C. (2016). The Hermeneutics of Trace Data: Building an Apparatus.

Østerlund, C., Crowston, K., & Jackson, C. (2019). Building an Apparatus: Refractive, Reflective & Diffractive Readings of Trace Data. *Journal of the Association for Information Systems (JAIS)*.

2) Yleistä tutkimuksen metodologiasta kuten subjektiivinen vai objektiivinen tieto, datalähtöinen vai mallilähtöinen tutkimusote

Braeken, J., Mulder, J., & Wood, S. (2015). Relative effects at work: Bayes factors for order hypotheses. *Journal of Management*, 41(2), 544-573.

Cox, T., Karanika, M., Griffiths, A., & Houdmont, J. (2007). Evaluating organizational-level work stress interventions: Beyond traditional methods. *Work & Stress*, 21(4), 348-362. doi:10.1080/02678370701760757

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *Academy of Management Review*, 14(4), 532-550. doi:10.5465/amr.1989.4308385

Eisenhardt, K. M., & Graebner, M. E. (2007). THEORY BUILDING FROM CASES: OPPORTUNITIES AND CHALLENGES. *Academy of Management Journal*, 50(1), 25-32. doi:10.5465/AMJ.2007.24160888

García-Herrero, S., Mariscal, M., Gutiérrez, J. M., & Ritzel, D. O. (2013). Using Bayesian networks to analyze occupational stress caused by work demands: Preventing stress through social support. *Accident Analysis & Prevention*, 57, 114-123.

Gehman, J., Glaser, V., Eisenhardt, K. M., Gioia, D. A., & Langley, A. (2016). Theory-Method Packages: A Comparison of Three Qualitative Approaches to Theory Building. *Academy of Management Proceedings*, 2016(1), 12424. doi:10.5465/ambpp.2016.12424symposium

Gehman, J., Glaser, V. L., Eisenhardt, K. M., Gioia, D., Langley, A., & Corley, K. G. (2018). Finding theory–method fit: A comparison of three qualitative approaches to theory building. *Journal of Management Inquiry*, 27(3), 284-300.

Goldstein, J. (1999). Emergence as a Construct: History and Issues. *Emergence*, 1(1), 49-72. doi:10.1207/s15327000em0101\_4

Hauser, O. P., Linos, E., & Rogers, T. (2017). Innovation with field experiments: Studying organizational behaviors in actual organizations. *Research in Organizational Behavior*, 37, 185-198. doi:https://doi.org/10.1016/j.riob.2017.10.004

Ilies, R., Aw, S. S., & Lim, V. K. (2016). A naturalistic multilevel framework for studying transient and chronic effects of psychosocial work stressors on employee health and well-being. *Applied Psychology*, 65(2), 223-258.

Langley, A., & Klag, M. (2017). Being Where? Navigating the Involvement Paradox in Qualitative Research Accounts. *Organizational Research Methods*, 22(2), 515-538. doi:10.1177/1094428117741967

Markus, M. L., & Robey, D. (1988). Information Technology and Organizational Change: Causal Structure in Theory and Research. *Management Science*, 34(5), 583-598. doi:10.1287/mnsc.34.5.583

Meyer, A. D., Tsui, A. S., & Hinings, C. R. (1993). Configurational Approaches to Organizational Analysis. *Academy of Management Journal*, 36(6), 1175-1195. doi:10.5465/256809

Schmitz, L. L., McCluney, C. L., Sonnega, A., & Hicken, M. T. (2019). Interpreting Subjective and Objective Measures of Job Resources: The Importance of Sociodemographic Context.

Spector, P. E., & Fox, S. (2003). Reducing subjectivity in the assessment of the job environment: Development of the Factual Autonomy Scale (FAS). *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 24(4), 417-432.

Spector, P. E. (2017). The lost art of discovery: The case for inductive methods in occupational health science and the broader organizational sciences. *Occupational health science*, 1(1-2), 11-27.

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2019). *Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field*: SAGE Publications Sage CA: Los Angeles, CA.

### 3) HR datan hyödyntämistä pohtivia

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: why HR is set to fail the big data challenge. *Human Resource Management Journal*, 26(1), 1-11. doi:10.1111/1748-8583.12090

Cappelli, P., Tambe, P., & Yakubovich, V. (2020). Can Data Science Change Human Resources? *The Future of Management in an AI World* (pp. 93-115): Springer.

Kovach, K. A., & Cathcart, C. E. (1999). Human Resource Information Systems (HRIS): Providing Business with Rapid Data Access, Information Exchange and Strategic Advantage. *Public Personnel Management*, 28(2), 275-282. doi:10.1177/009102609902800208

McAbee, S. T., Landis, R. S., & Burke, M. I. (2017). Inductive reasoning: The promise of big data. *Human Resource Management Review*, 27(2), 277-290.

Rasmussen, T., & Ulrich, D. (2015). Learning from practice: how HR analytics avoids being a management fad. *Organizational Dynamics*, 44(3), 236-242. doi:10.1016/j.orgdyn.2015.05.008





Shah, N., Irani, Z., & Sharif, A. M. (2017). Big data in an HR context: Exploring organizational change readiness, employee attitudes and behaviors. *Journal of Business Research*, 70, 366-378.

Sheng, J., Amankwah-Amoah, J., & Wang, X. (2017). A multidisciplinary perspective of big data in management research. *International Journal of Production Economics*, 191, 97-112.

van der Laken, P. (2018). Data-driven human resource management: The rise of people analytics and its application to expatriate management. Ridderkerk: Ridderprint.

#### 4) Ehdotuksia datatieteilijöiden näkökulmasta työdatan hyödyntämisestä

Braun, M. T., Kuljanin, G., & DeShon, R. P. (2018). Special considerations for the acquisition and wrangling of big data. *Organizational Research Methods*, 21(3), 633-659.

Calvard, T. S. (2016). Big data, organizational learning, and sensemaking: Theorizing interpretive challenges under conditions of dynamic complexity. *Management Learning*, 47(1), 65-82.

Dunsmuir, W. T., & Scott, D. J. (2015). The glarma package for observation-driven time series regression of counts. *Journal of Statistical Software*, 67(7), 1-36.

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62-70. doi:<https://doi.org/10.1016/j.infoandorg.2018.02.005>

Gaskin, J. E., Berente, N., Lyytinen, K., & Yoo, Y. (2014). Toward Generalizable Sociomaterial Inquiry: A Computational Approach for Zooming In and Out of Sociomaterial Routines. *Mis Quarterly*, 38(3), 849-871.

Ghosh, A., Tweedale, J. W., & Nafalski, A. (2013). Modified hybridized multi-agent oriented approach to analyze work-stress data providing feedback in real time. *Procedia Computer Science*, 22, 1092-1101.

Hedman, J., Srinivasan, N., & Lindgren, R. (2013). Digital traces of information systems: Sociomateriality made researchable.

Johnson, S. L., Gray, P., & Sarker, S. (2019). Revisiting IS research practice in the era of big data. *Information and Organization*, 29(1), 41-56. doi:[10.1016/j.infoandorg.2019.01.001](https://doi.org/10.1016/j.infoandorg.2019.01.001)

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 2053951714528481. doi:[10.1177/2053951714528481](https://doi.org/10.1177/2053951714528481)

- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 5802-5805. (viittaus trace dataan)
- Kowatsch, T., Wahle, F., Filler, A., Kehr, F., Volland, D., Haug, S., Fleisch, E. (2015). Towards short-term detection of job strain in knowledge workers with a minimal-invasive information system service: Theoretical foundation and experimental design.
- Kowatsch, T., Wahle, F., & Filler, A. (2017). stressOUT: Design, Implementation and Evaluation of a Mouse-based Stress Management Service Designing the Digital Transformation: DESRIST 2017 Research in Progress Proceedings (pp. 37-45).
- Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers. Paper presented at the Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Republic of Korea.
- Luciano, M. M., Mathieu, J. E., Park, S., & Tannenbaum, S. I. (2017). A Fitting Approach to Construct and Measurement Alignment: The Role of Big Data in Advancing Dynamic Theories. *Organizational Research Methods*, 21(3), 592-632. doi:10.1177/1094428117728372
- Menger, V., Spruit, M., Hagoort, K., & Scheepers, F. (2016). Transitioning to a data driven mental health practice: collaborative expert sessions for knowledge and hypothesis finding. *Computational and mathematical methods in medicine*, 2016.
- Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital Innovation Management: Reinventing innovation management research in a digital world. *Mis Quarterly*, 41(1). (viittaus digital trace dataan)
- Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'. *The Journal of Strategic Information Systems*, 24(1), 3-14. doi:<https://doi.org/10.1016/j.jsis.2015.02.001>
- Pentland, B. T., Pentland, A. P., & Calantone, R. J. (2017). Bracketing off the actors: Towards an action-centric research agenda. *Information and Organization*, 27(3), 137-143. (tässä jutussa ehdottavat Engeströmin mallia datalähtöisen systeemin tutkimuksessa)
- Pentland, B., Recker, J., & Kim, I. (2017). Capturing reality in flight? Empirical tools for strong process theory.
- Schildt, H. (2017). Big data and organizational design—the brave new world of algorithmic management and computer augmented transparency. *Innovation*, 19(1), 23-30.

Suriadi, S., Wynn, M. T., Xu, J., van der Aalst, W. M. P., & ter Hofstede, A. H. M. (2017). Discovering work prioritisation patterns from event logs. *Decision Support Systems*, 100, 77-92. doi:<https://doi.org/10.1016/j.dss.2017.02.002>

Tonidandel, S., King, E. B., & Cortina, J. M. (2016). Big Data Methods: Leveraging Modern Data Analytic Techniques to Build Organizational Science. *Organizational Research Methods*, 21(3), 525-547. doi:10.1177/1094428116677299

Tweedale, J. W. (2016). Decision-making in a distributed and dynamically scalable environments Intelligent decision technology support in practice (pp. 107-124): Springer.

5) Organisaatiodynamiikka, käytäntöjä, prosesseja ja aikaa sekä digital trace dataa käsitteleviä artikkeleita, joita voisi soveltaa työn aikasarja-mallintamiseen

#### a) Käytännöt

Barley, S. R., & Kunda, G. (2001). Bringing work back in. *Organization science*, 12(1), 76-95

Feldman, M. S., & Orlikowski, W. J. (2011). Theorizing Practice and Practicing Theory. *Organization Science*, 22(5), 1240-1253. doi:10.1287/orsc.1100.0612

Feldman, M. S., Pentland, B. T., D'Adderio, L., & Lazaric, N. (2016). Beyond routines as things: introduction to the special issue on routine dynamics. *Organization Science*, 27(3), 505-513. (viittaus digital trace dataan)

Frese, M., & Zapf, D. (1994). Action as the core of work psychology: A German approach *Handbook of industrial and organizational psychology*, Vol. 4, 2nd ed. (pp. 271-340). Palo Alto, CA, US: Consulting Psychologists Press.

Goh, K. T., & Pentland, B. From Actions to Paths to Patterning: Toward a Dynamic Theory of Patterning in Routines. *Academy of Management Journal*(ja). (viittaus digital trace dataan)

Hacker, W. (2003). Action Regulation Theory: A practical tool for the design of modern work processes? *European Journal of Work and Organizational Psychology*, 12(2), 105-130. doi:10.1080/13594320344000075

Jennings, D. TOPIC MODELING IN MANAGEMENT RESEARCH: RENDERING NEW THEORY FROM TEXTUAL DATA.

Karasek, R. A. (1979). Job demands, job decision latitude and mental strain: Implications for job redesign. *Administrative Science Quarterly*, 24(2), 285-308. doi.org/10.2307/2392498

#### b) Prosessit

Abdallah, C., Lusiani, M., & Langley, A. (2019). Performing process research Standing on the Shoulders of Giants: Traditions and Innovations in Research Methodology (pp. 91-113): Emerald Publishing Limited.

Binsch, O., Wabeke, T., & Valk, P. (2016). Comparison of three different physiological wrist-band sensor systems and their applicability for resilience-and work load monitoring. Paper presented at the 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN).

Cheng, M. M., Li, C., Hackett, R. D., & Lee-Chin, M. (2018). Simulation and big data: in search of causality in big data-related managerial decision making. (viittaus digital trace dataan)

Choi, S. (2018). When digital trace data meet traditional communication theory: Theoretical/Methodological Directions. *Social Science Computer Review*, 0894439318788618. (viittaus digital trace dataan)

Dittrich, K., Feldman, M. S., & Pentland, B. T. PDW 01: Making Process Visible: Recognizing and Visualizing Action Patterns. (viittaus digital trace dataan)

Eckardt, R., Crocker, A., Ahn, Y., Floyd, S. W., Boyd, B. K., Hodgkinson, G. P., Starbuck, W. H. (2018). Reflections on the micro-macro divide: Ideas from the trenches and moving forward. *Strategic Organization*, 1476127018770048.

Kouamé, S., & Langley, A. (2018b). Relating microprocesses to macro-outcomes in qualitative strategy process and practice research. *Strategic Management Journal*, 39(3), 559-581.

Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management review*, 24(4), 691-710.

Langley, A., Smallman, C., Tsoukas, H., & Ven, A. H. V. d. (2013). Process Studies of Change in Organization and Management: Unveiling Temporality, Activity, and Flow. *Academy of Management Journal*, 56(1), 1-13. doi:10.5465/amj.2013.4001

Langley, A. (2019). Visual Artifacts as Tools for Analysis and Theorizing. In D. Ravasi, B. Z. Tammar, M. A. John, & M. Johanna (Eds.), *The Production of Managerial Knowledge and Organizational Theory: New Approaches to Writing, Producing and Consuming Theory* (Vol. 59, pp. 173-199): Emerald Publishing Limited.

Leenders, R. T. A., Contractor, N. S., & DeChurch, L. A. (2016). Once upon a time: Understanding team processes as relational event networks. *Organizational Psychology Review*, 6(1), 92-115. (digital time stamped data)

McIntyre, T. M., McIntyre, S. E., Barr, C. D., Woodward, P. S., Francis, D. J., Durand, A. C., Kamarck, T. W. (2016). Longitudinal study of the feasibility of using ecological momentary

assessment to study teacher stress: Objective and self-reported measures. *Journal of Occupational Health Psychology*, 21(4), 403.

Rosa, R. A., & Bulgacov, S. Sub-theme 03: [SWG] Routines, Stability and Change in Organizations and Beyond. (viittaus digital trace dataan)

Spector, P. E., & Meier, L. L. (2014). Methodologies for the study of organizational behavior processes: How to find your keys in the dark. *Journal of Organizational Behavior*, 35(8), 1109-1119.

Spector, P. E., & Pindek, S. (2016). The future of research methods in work and occupational health psychology. *Applied Psychology*, 65(2), 412-431. (viittaus big dataan)

Suh, C., & Lee, H. (2018). Hidden footage in online communities: The assimilation processes of newcomers under the fluidity. (viittaus digital trace dataan)

Weick, K. E., & Quinn, R. E. (1999). ORGANIZATIONAL CHANGE AND DEVELOPMENT. *Annual Review of Psychology*, 50(1), 361-386. doi:10.1146/annurev.psych.50.1.361

### c) Aika

Ancona, D. G., Okhuysen, G. A., & Perlow, L. A. (2001). Taking time to integrate temporal research. *Academy of Management Review*, 26(4), 512-529.

Cesare, N., Lee, H., McCormick, T., Spiro, E., & Zagheni, E. (2018). Promises and pitfalls of using digital traces for demographic research. *Demography*, 55(5), 1979-1999.

Navarro Cid, J., Roe, R. A., & Artiles, M. I. (2015). Taking time seriously: Changing practices and perspectives in Work/Organizational Psychology. *Revista de Psicología del Trabajo y de las Organizaciones= Journal of Work and Organizational Psychology*, 2015, vol. 31, p. 135-145.

Sonnentag, S. (2012). Time in organizational research: Catching up on a long neglected topic in order to improve theory. *Organizational Psychology Review*, 2(4), 361-368.

Spears, D. R., McNeil, C., Warnock, E., Trapp, J., Oyinloye, O., Whitehurst, V., Meechan, P. (2013). Predicting temporal trends in sickness absence rates for civil service employees of a federal public health agency. *Journal of occupational and environmental medicine*, 55(2), 179-190.

Vanttilborgh, T., Hofmans, J., & Judge, T. A. (2018). The time has come to study dynamics at work. *Journal of Organizational Behavior*, 39(9), 1045-1049.

### 6) Asiantuntijan työaikaa organisaatiotason ilmiönä käsittelevät tutkimukset

Blagoev, B., Muhr, S. L., Ortlieb, R., & Schreyögg, G. (2018). Organizational working time regimes: Drivers, consequences and attempts to change patterns of excessive working

hours. *German Journal of Human Resource Management*, 32(3-4), 155-167. doi:10.1177/2397002218791408

Costas, J., Ekman, S., Empson, L., Kärreman, D., & Muhr, S. L. (2018). Working time regimes: A panel discussion on continuing problems. *German Journal of Human Resource Management*, 32(3-4), 271-282. doi:10.1177/2397002218791389

Decuyper, M., & Simons, M. (2014). On the Composition of Academic Work in Digital Times. *European Educational Research Journal*, 13(1), 89-106. doi:10.2304/eeerj.2014.13.1.89

Elfering, A., Grebner, S., Ganster, D. C., Berset, M., Kottwitz, M. U., & Semmer, N. K. (2018). Cortisol on Sunday as indicator of recovery from work: Prediction by observer ratings of job demands and control. *Work & Stress*, 32(2), 168-188.

Mullan, K., & Wajcman, J. (2017). Have Mobile Devices Changed Working Patterns in the 21st Century? A Time-diary Analysis of Work Extension in the UK. *Work, Employment and Society*, 33(1), 3-20. doi:10.1177/0950017017730529

Perlow, L. A. (1999). The Time Famine: Toward a Sociology of Work Time. *Administrative Science Quarterly*, 44(1), 57-81. doi:10.2307/2667031

Wajcman, J. (2018a). The Digital Architecture of Time Management. *Science, Technology, & Human Values*, 44(2), 315-337. doi:10.1177/0162243918795041

Wajcman, J. (2018b). Digital technology, work extension and the acceleration society. *German Journal of Human Resource Management*, 32(3-4), 168-176. doi:10.1177/2397002218775930

Wajcman, J. (2018c). How Silicon Valley sets time. *New Media & Society*, 21(6), 1272-1289. doi:10.1177/1461444818820073

7) Moodle-dataa, "virtual learning environment" -dataa, hyödyntävät "learning analytics" teemaiset tutkimukset, joissa ennakoitiin *opiskelijan* edistymistä.

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational Data Mining and Learning Analytics for 21st century higher education: A Review and Synthesis. *Telematics and Informatics*.

Daniel, B. K. (2019). Big Data and data science: A critical review of issues for educational research. *British Journal of Educational Technology*, 50(1), 101-113.

Drlik, M., & Munk, M. (2018). Understanding Time-Based Trends in Stakeholders' Choice of Learning Activity Type Using Predictive Models. *IEEE Access*, 7, 3106-3121.

Drlik, M. (2018). Understanding of Time-Based Trends in Virtual Learning Environment Stakeholders' Behaviour. Paper presented at the International Conference on Interactive Collaborative Learning.

Felix, I., Ambrosio, A. P., Duilio, J., & Simões, E. Predicting Student Outcome in Moodle.

Howard, E., Meehan, M., & Parnell, A. (2018). Contrasting prediction methods for early warning systems at undergraduate level. *The Internet and Higher Education*, 37, 66-75.

Lim, Y. M. (2017). Detecting and Modelling Stress Levels in E-Learning Environment Users.

Młynarska, E., Greene, D., & Cunningham, P. (2016). Time series clustering of Moodle activity data. Paper presented at the 24th Irish Conference on Artificial Intelligence and Cognitive Science (AICS'16), University College Dublin, Dublin, Ireland, 20-21 September 2016.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39.

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *Internet and Higher Education*, 29, 23–30.

Moodlen käyttö on maailmanlaajuista: <https://stats.moodle.org/>

.....

## LDA

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.

Hoffman, M., Bach, F. R., & Blei, D. M. (2010). Online learning for latent dirichlet allocation. In *advances in neural information processing systems* (pp. 856-864).



## LIITTEET

### Liite A: Käytäntöjen toteutumisen aste viikoittain

viikko	Käytäntö																
	#0	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16
2017																	
1	2	18	3	6	6	12	1	5	2	1	9	1	0	11	3	1	2
2	4	42	7	13	20	16	6	8	4	4	39	1	0	38	10	3	10
3	3	29	9	11	19	9	5	12	2	3	42	2	0	41	8	5	11
4	2	32	9	9	21	9	6	12	2	3	37	1	0	39	6	5	7
5	2	32	9	9	21	6	4	12	3	3	47	1	1	44	7	4	9
6	2	29	6	8	21	8	5	9	2	3	39	0	0	34	7	2	8
7	2	31	7	7	22	7	5	8	2	3	45	2	0	36	5	3	7
8	2	27	6	6	18	9	5	9	3	2	40	1	0	36	5	5	8
9	1	9	4	2	6	2	1	4	0	3	13	0	0	8	1	2	1
10	2	35	8	7	19	10	4	9	4	6	43	2	0	41	6	6	8
11	2	33	6	8	21	8	5	8	2	4	38	2	0	32	6	3	6
12	2	30	8	9	20	7	6	10	3	5	36	2	1	40	3	3	7
13	2	31	9	8	17	5	4	13	1	4	44	2	0	38	3	4	8
14	3	27	10	6	20	5	3	15	1	5	37	2	0	39	2	6	6
15	1	22	8	6	18	4	2	12	1	4	33	1	0	30	2	3	5
16	2	21	8	6	21	4	3	12	1	4	43	1	0	37	2	4	7
17	2	20	8	4	22	7	3	14	3	3	36	1	0	42	2	6	7
18	3	12	12	4	16	3	1	14	2	4	33	1	0	50	5	4	7
19	2	12	9	3	14	5	3	13	2	4	26	0	0	51	3	4	8
20	2	12	10	6	16	5	2	13	2	5	19	1	0	46	4	4	9
21	2	9	13	2	12	4	1	17	2	6	16	1	1	41	1	5	5
22	2	10	9	2	10	8	2	14	1	4	15	1	0	42	2	2	6
23	1	12	8	3	11	10	2	9	2	4	13	0	0	42	4	5	5
24	1	5	3	1	5	3	0	5	1	1	5	1	0	13	1	2	3
25	1	2	2	0	6	1	0	3	0	1	4	0	0	9	0	0	1
26	0	2	1	0	2	1	0	0	0	1	3	0	2	3	1	0	0
27	0	0	2	0	2	1	0	1	0	0	1	0	2	3	0	0	0
28	0	1	1	0	2	0	0	1	0	0	1	0	1	3	0	0	0
29	0	0	1	0	2	0	0	1	1	0	2	0	1	1	0	0	0
30	0	1	3	0	1	2	0	1	0	0	1	0	1	2	0	0	0
31	0	3	3	1	3	2	0	1	1	1	2	0	7	8	1	0	1
32	1	15	8	4	7	11	1	3	2	3	5	1	27	21	4	1	4
33	1	21	5	5	9	24	2	2	3	3	12	2	35	18	6	1	5
34	2	35	6	9	11	29	3	3	3	9	20	2	45	20	7	2	9
35	3	37	10	9	15	17	4	5	3	4	24	2	56	25	7	3	10
36	3	32	11	12	17	11	5	6	2	4	25	2	54	29	8	3	10





37	2	30	8	11	20	13	3	4	4	5	36	2	49	25	6	3	7
38	3	29	13	8	22	11	3	7	2	6	29	2	47	27	4	3	8
39	2	23	16	7	20	9	3	8	2	4	36	2	40	29	4	3	4
40	4	30	14	7	19	9	2	9	1	4	37	1	45	28	4	5	7
41	2	25	14	5	18	10	2	8	1	6	34	2	39	28	6	5	6
42	1	9	7	3	9	5	1	4	1	2	8	2	13	14	2	3	1
43	3	25	15	7	18	11	2	10	2	6	37	2	42	29	5	5	8
44	3	27	16	8	17	10	2	10	2	2	30	2	45	30	6	4	9
45	4	21	18	7	19	7	3	12	1	3	28	1	43	33	5	5	8
46	6	19	22	9	19	7	3	14	1	3	39	3	45	35	2	6	7
47	3	19	18	7	19	5	2	12	2	3	36	1	39	35	4	5	6
48	4	19	17	6	17	5	2	11	2	4	27	3	39	37	2	5	7
49	5	20	20	6	17	7	2	11	1	5	22	1	36	40	4	6	7
50	5	20	18	6	14	7	2	12	2	3	25	2	39	40	3	5	6
51	5	17	23	3	13	7	1	14	1	3	15	1	29	38	5	4	7
52	0	1	3	0	1	1	0	5	0	0	0	0	2	9	0	0	1
2018																	
1	1	14	5	2	5	6	2	4	1	1	3	1	11	11	2	1	3
2	6	36	13	12	11	16	3	8	3	4	21	3	53	29	11	4	10
3	4	24	19	8	14	8	4	10	2	3	28	2	43	31	9	4	8
4	4	24	22	8	15	7	3	11	1	4	28	2	42	34	6	3	8
5	5	21	21	8	14	8	2	11	3	4	27	2	40	40	5	4	7
6	5	26	18	7	17	9	2	9	1	3	27	3	41	32	5	4	7
7	5	25	16	8	16	8	3	9	2	4	22	2	38	28	3	4	4
8	5	22	16	6	15	9	1	9	1	5	23	2	38	33	6	5	5
9	2	10	6	1	10	3	1	4	0	2	9	0	11	11	3	5	2
10	5	26	16	7	19	12	2	9	1	8	26	3	46	36	8	7	6

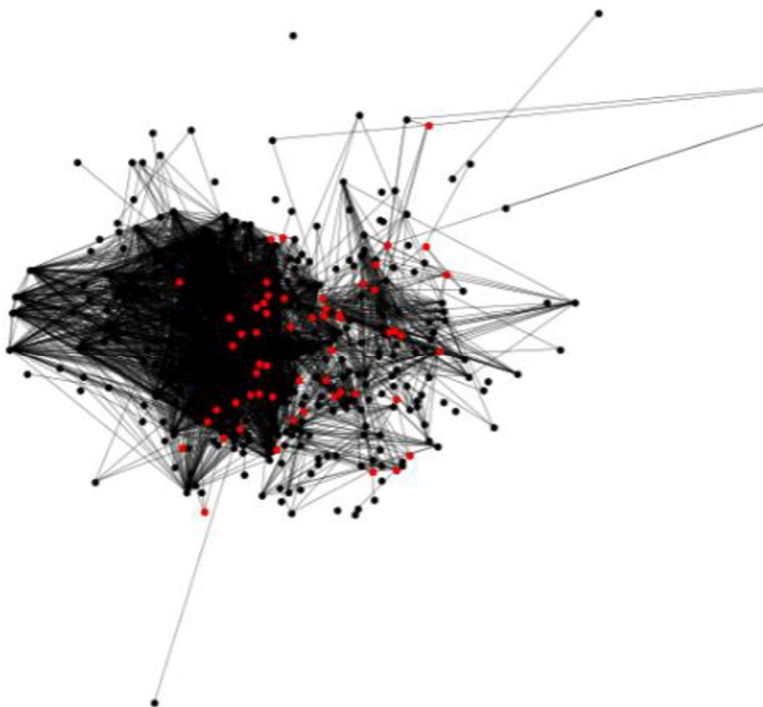
## Liite B: Verkostoanalyysi

Yhteistyön asteen aikasarja saatiin muodostamalla datan avulla ensin viikkokohtaisesti suuntaamaton verkko. Verkossa opettajat ovat solmuja ja solmujen välillä oleva kaari merkitsee, että kyseisillä opettajilla on ainakin yksi yhteinen tehtävä/tapahtuma kyseisellä viikolla. Solmun aste on kustakin solmusta (opettajasta) lähtevien kaarten lukumäärä. Tässä tapauksessa solmun aste tarkoittaa, kuinka monen eri opettajan kanssa tällä opettajalla oli suunniteltua tekemistä kyseisellä viikolla. Itse aikasarja muodostui kunkin viikon eri opettajien yhteistyön asteiden summasta.

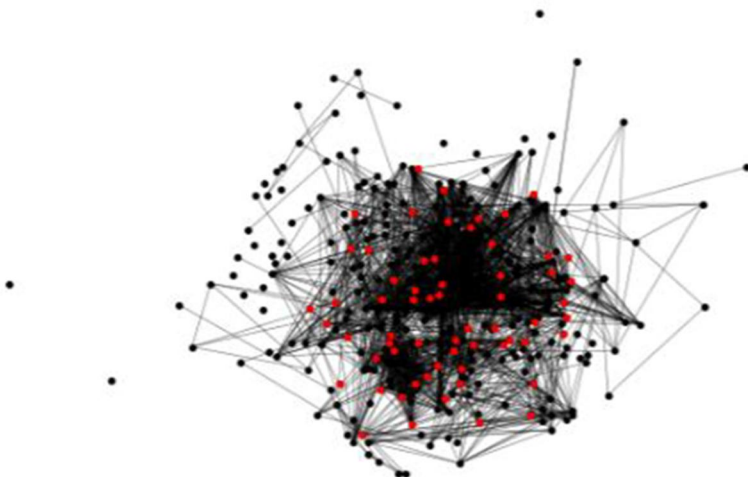
Kyse ei siis ole vapaasti kehkeytyvästä verkostosta vaan verkostosta, jossa henkilöt suunnitellusti kiinnitetään yhteisiin tapahtumiin/tehtäviin.

Esimerkkikuvat ovat viikoilta (2017) 14, 35 ja 46. Kaaret kertovat opettajan yhteistyökumppaneista ja punaiset pisteet kertovat keskeisyydestä, jota tietoa ei tässä raportissa tulkittu.

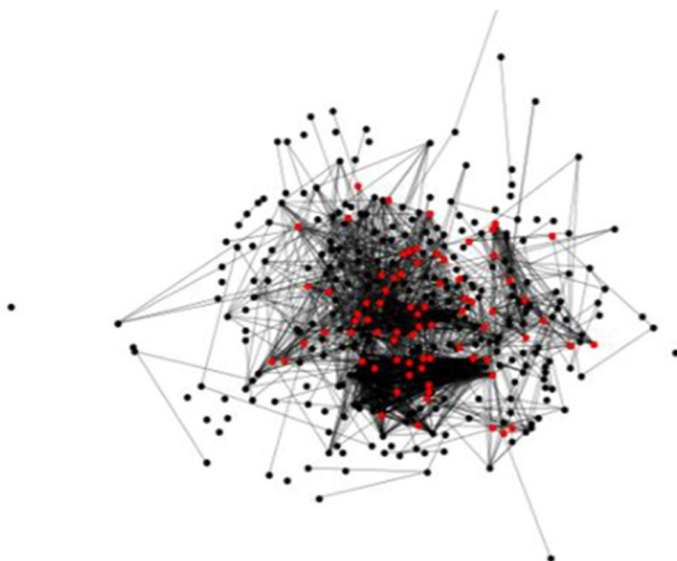
### Viikko 14



Viikko 35



Viikko 46



## LIITE C: Hankkeen aikana (2018-2019) kirjoitetut kuusi blogit

### 1) Työdatan hiljaiset signaalit: Mahdollisuuksien meri

Tämä kirjoitus aloittaa blogien sarjan, jossa pohdimme, miten hyödynnämme työdataa työolojen kehittämässä. Työdatalla tarkoitamme tapahtumatietoja, joita jatkuvasti ja tu-  
levaisuudessa enenevässä määrin kertyy käyttämiimme digitaalisiin järjestelmiin.

<https://www.ttl.fi/blogi/tyhis-mahdollisuuksien-meri/>

### 2) Mitä on datalähtöinen työn ja käytäntöjen tutkimus?

Työelämä tutkimuksessa saadaan yhä enemmän rikasta dataa, jota kertyy järjestelmiin il-  
man erillistä tietojen keräystä. Tässä TYHIS-blogisarjan osassa kuvataan, miten dataa  
lähestytään tutkimuksessa. Miten siitä saadaan digitaalisia jalanjälkiä sekä massadataa?

<https://www.ttl.fi/blogi/mita-on-datalahtoinen-tyon-ja-kaytantojen-tutkimus/>

### 3) Tekoäly ja uudet aineistolähteet avuksi muutoksen hallintaan

Millaista on tekoälyn hyödyntäminen työelämä tutkimuksessa? Tässä TYHIS-blogisarjan  
osassa kuvataan konkreettisesti, kuinka tutkijat louhivat opettajien digitaalisista jalanjäl-  
jistä tietoa viikkokäytännöistä. Ja ne voivat kertoa työhyvinvoinnin kannalta tärkeää sano-  
maa.

<https://www.ttl.fi/blogi/tekoaly-ja-uudet-aineistolahteet-avuksi-muutoksen-hallintaan/>

### 4) Mitä digitaalinen jalanjälki kertoo opettajien arki- ja sunnuntaityöstä vuoden ai- kana?

Digitaalinen jalanjälki paljastaa organisaation työn rytmin. Sen avulla voidaan tunnistaa  
erilaista työn määrän vaihtelua viikosta toiseen. Viikoittainen työn rytmi tarjoaa uuden  
näkökulman toiminnan suunnitteluun ja johtamiseen.

<https://www.ttl.fi/blogi/mita-digitaalinen-jalanjalki-kertoo-opettajien-arki-ja-sunnuntai-tyosta-vuoden-aikana/>

### 5) Työt ja poissaolot kulkevat samassa viikkotahdissa – tekoäly auttaa analyysissä Ammattikorkeakoulun syksy alkaa tasaisesti lisääntyvällä kurssien parissa työskentelyllä ja samaa tahtiin lisääntyvät lyhyet sairauspoissaolot. Tekoäly seuloa tietomassaa, tutkija pohtii syitä.

<https://www.ttl.fi/blogi/tyot-ja-poissaolot-kulkevat-samassa-viikkotahdissa-tekoaly-auttaa-analyysissa/>



## 6) SILENT SIGNALS FROM WORK DATA: PROFILING CONDITIONS, NOT INDIVIDUALS

Is it possible to utilize digital footprint of work, work data, to design better working conditions? We define work data as event traces, event logs, which gathered in IT-systems continuously and in the future even more.

<https://www.ttl.fi/en/silent-signals-from-work-data-profiling-conditions-not-individuals/>

Hankkeen nettisivut: [www.ttl.fi/tyhis](http://www.ttl.fi/tyhis)

Raportissa kerrotaan ”Työdatan hiljaiset signaalit” -hankkeen vaiheista, joilla työelämä tutkimuksen uusi aineistolähde ”digital trace data”, työdata, saatiin käytettävään muotoon. Aikaleimattua dataa kerääntyy jatkuvasti ja automaattisesti työssä käytettyjen järjestelmien lokitietoihin. Työdataa tarkastella koko organisaation yhteisen tekemisen näkökulmasta.

Hankkeen työdata on yhden ammattikorkeakoulun Moodlen lokitietoihin jäävät käyttötapahtumat 63 viikolta. Tapahtumia tarkasteltiin tekstinä ja käytäntöinä. Tekstiaineistoa analysoivana tekoälynä sovellettiin LDA -aihemallinnusta, joka tunnisti 17 käytäntöä. Viikkotason tarkastelu tehtiin käytäntöjen lisäksi sairauspoissaolo- ja suunnittelujärjestelmistä saaduista tiedoista. Esittelemme viikkorytmien analyysitavan ja periaatteet, joita voidaan soveltaa moniin vastaaviin aineistolähteisiin.

Käytäntöjen ja työhyvinvoinnin kehittämistä on nyt mahdollista toteuttaa uudella tavalla. Jos tulevaisuudessa viikkorytmien seuranta on automaattista, voi kehittämisen tulosten seurantakin olla jatkuvaa.



Työsuojelurahasto  
Arbetskyddsfonden  
The Finnish Work Environment Fund

Työterveyslaitos  
Arbetshälsainstitutet  
Finnish Institute of Occupational Health

PL 40, 00032 Työterveyslaitos

[www.ttl.fi](http://www.ttl.fi)

ISBN 978-952-261-909-9 (nid.)

ISBN 978-952-261-908-2 (PDF)