



**Kaakkois-Suomen  
ammattikorkeakoulu**



**South-Eastern Finland  
University of Applied Sciences**

## **HUOM! TÄMÄ ON RINNAKKAISTALLENNE**

**Rinnakkaistallennettu versio voi erota alkuperäisestä julkaistusta sivunumeroiltaan ja ilmeeltään.**

**Tekijä(t): Jääskeläinen, Anssi**

**Otsikko: DLM Triennial konferenssi Brightonissa**

**Versio: julkaisijan versio/PDF**

**Käytä viittauksessa alkuperäistä lähdettä:**

**Jääskeläinen, A., 2017. DLM Triennial konferenssi Brightonissa. Faili 4/2017.**

## **PLEASE NOTE! THIS IS PARALLEL PUBLISHED VERSION / SELF-ARCHIVED VERSION OF THE OF THE ORIGINAL ARTICLE**

**This is an electronic reprint of the original article.**

**This version may differ from the original in pagination and typographic detail.**

**Author(s): Jääskeläinen, Anssi**

**Title: DLM Triennial konferenssi Brightonissa. Faili 4/2017.**

**Version: publisher's version/PDF**

**Please cite the original version:**

**Jääskeläinen, A., 2017. DLM Triennial konferenssi Brightonissa. Faili 4/2017.**

## DLM Triennial

## konferenssi Brightonissa



Anssi  
Jääskeläinen  
TKI-asiantuntija  
Digitalia/Xamk

Tällä kertaa DLM:n jäsenkokouksen tilalla oli kerran kolmessa vuodessa järjestettävä Triennial konferenssi, jonka pitopaikkana oli Brighton. Matka alkoi "hyvin" lennon ollessa jo lähtiessä melkein kolme tuntia myöhässä Ranskassa olleen lennonjohtajien lakon vuoksi. Allekirjoittaneella oli matkatovereina kollegat *Mikko Lampi* ja *Noora Talsi*. Xamkin edustajistolla oli konferenssissa kaksi paperia esitettävänä, mutta siitäkin huolimatta ohjelma vaikutti kiinnostavalta. Luvassa oli muun muassa käytännöllistä tietokantojen arkistointia, lohkoketjutekniikkaan perustuvaa säilyttämistä sekä E-ark-projektin mahdolliset jatkokuulumiset.

Uusi puheenjohtaja, Tanskan kansallisarkiston *Jan Dalsten Sørensen*, avasi konferenssin perinteiseen tapaan kiitoksilla paikasta ja järjestelyistä. Tästä jatkoi *Chris Pole*, Brightonin yliopiston varakansleri. Hän kertoi yleisiä asioita yliopistosta, jossa on kaikkiaan yli 22 000 opiskelijaa. Kampukset sijaitsevat Brightonissa, EastBournessissa ja Hastingsissa. Viimeisenä hän kiitti professori *Janet Andersonia* (aiemmin Delve) sekä

E-ark-projektia ja sen tuloksia koko Euroopan mittakaavassa.

### Tietokantojen arkistointi

Seuraavaksi Xamkin delegaatio osallistui Keep Solutionin pitämään tietokantojen arkistointityöpajaan – tosin kahdella oli kahvitauosta eteenpäin tärkeämpää tekemistä. Allekirjoittaneelle ei tullut yllätyksenä, että tietokantojen arkistointi perustui Siard 2.0:aan. Niille lukijoille, jotka eivät Siardia vielä kukaan tunne, kerrotakoon, että se on tapa, jolla tietokanta voidaan muuttaa yleisluontoiseen muotoon, joka on helposti siirrettävissä eri tietokantaohjelmien välillä. Tällä hetkellä Siard tukee käytännössä kaikkia yleisesti käytössä olevia tietokantoja. Teknisesti Siard on zip-paketti, joka pitää sisällään kansiorakenteen, jossa on otsikko- ja sisältötiedot. Tämän voi todentaa helposti muuttamalla .siard-päätteen .zip-päätteeksi ja purkamalla rakenteen. Siard-formaatista on kehitteillä myös uusi versio, mutta taaksepäin yhteensopivuutta Siard 2.0-pakettien kanssa ei uskallettu luvata.

Varsinaisen konferenssiohjelman aloitti *Nina Bryant*, Information Squared Consulting Ltd. Hän puhui arkistoista ja tietojärvistä sekä niiden mahdollisesta yhdistymisestä. Puheen lähtökohtana oli arkiston ja varmuuskopion ero. Puheessa sivuttiin myös monia muita arkistoinnin kannalta tärkeitä asioita kuten verkkolevyillä olevaa raakada-

taa, vanhojen arkistojärjestelmien tekoheittämistä sekä johtoportaan kehnoa ymmärrystä arkistoinnista. Johtoportaan mukaan yli 80% käyttäjistä luulee, että arkistointia tehdään lainsäädännön pakon takia. Ninan mielestä arkistoinnin pitäisi kuitenkin säilyttää ja saattaa käyttöön tietoa, joka hyödyttää tulevaisuuden päätöksenteossa ja edesauttaa muun muassa tunnistamaan uusia potentiaalisia asiakasryhmiä.

Seuraavan monotonisen lukuesityksen ainoa järjevä asia oli toteamus, että tarvittaisiin yleisesti hyväksytyt metatieto- ja siirtoformaatit, jotka mahdollistaisivat ERMS (Electronic Records Management System) järjestelmien saumattoman yhteentömmivuuden ja tietojen helpon siirrettävyyden. Käsitteäkseni tähän on juuri se asia, johon Moreq on tähännyt syntymästään asti.

### Palvelut ja sääntely Virossa

Tuttavamme *Liivi Karpistsenko* Viron talous- ja viestintäministeriöstä kertoi palveluista ja sääntelystä Virossa. Suuri osa hallinnollisesta tiedosta siirtyy sähköisenä X-roadin kautta ja tästäkin määrästä 75% kulkee tietona eikä siis dokumentteina. Viron kansalaisten sähköinen ID toimii jokaisessa järjestelmässä. Peräti 99% palveluista on sähköisiä ja virolaiset luottavat niihin. Mielenkiintoisena yksityiskohtana mainittakoon, että Virossa ei ole tarpeen tarjota palvelua paperilla vaan kansalaisilta





Brightonin yliopisto, Cockerst Building.

edellytetään sähköisien palveluiden käyttöä. Miksi tätä ei voisi tapahtua myös Suomessa? Me täällä Suomessa, myös Xamkillä, tulostamme, allekirjoitamme ja skannaamme, jotta dokumentti on varmasti ”virallinen”. Sivuhuomautuksena voinee todeta, että kone aikansa opeteltuaan pystyy väärentämään henkilön kuin henkilön allekirjoituksen siten, ettei edes henkilö itse tunnista eroa.

Virossa uusimman voimaan tulleen lain ”Principles for Managing Services and Governing Information” 03/06/2017 avulla mennään vieläkin lähemmäs digitaalista yhteiskuntaa seuraavien ohjenuorien avulla:

1. Käyttäjät edellä: palvelut ja tiedot sen mukaan, mitä käyttäjät tarvitsevat.
2. Kertatallennus: järjestelmät keskustelevat keskenään ja hakevat tarvittavat tiedot sieltä, missä ne ovat. Pällekkäisiä tietoja vältetään viimeiseen asti.

3. MyData-tyyppinen käyttöoikeuksien hallinta/kontrolli, jossa henkilöllä itsellään on näkymä siihen, mitä tietoa viranomaisilla hänestä on ja kuka niitä on mihinkin tarkoitukseen käyttänyt.

### Standardeista

Iltapäivän viimeisessä sessiossa kerrottiin standardeista sekä niiden voimasta sääntelyssä ja vaikuttamisessa. Todettiin muun muassa standardeilla haettavan yhteensopivuutta, mutta kysyttiin minkä kanssa. Yhtenä esimerkkinä *Alan Shipman* (Group 5 Training Limited) nosti esiin ISO/IEC 27018:2014 standardin. Jos esimerkiksi pilvipalvelun tuottaja/tarjoaja väittää noudattavansa kyseistä standardia, tulee sen myös avoimesti kertoa, missä pilvi fyysisesti sijaitsee. Lisäksi palveluntarjoajan on informoitava käyttäjiään, jos pilven paikka muuttuu. Moniko näin oikeasti tekee?

Toisena esimerkkinä *Shipman* mainitsi ISO standardit yleisesti. Jos jollakin standardilla on vain ISO merkintä, se voi tarkoittaa, että kyseisessä maassa tai koko EU:ssa ei pidetä siitä. Sen sijaan BS (British Standard) EN (European Norm) ISO merkinnöillä varustettu standardi on oikeasti hyväksytty maailmanlaajuisesti, Suomen tapauksessa merkintä luonnollisesti on SFS-EN-ISO.

### Vahingollinen sähköinen arkisto

Torstain aloitti *David Anderson* kertomalla päivän ohjelmasta. Esitykset videoitiin ainakin tämän konferenssin ajan, mutta Anderson toivoi vastaavaa myös tulevaisuudessa, mukaan lukien jäsenkokoukset. Lisäksi pyritään mahdollistamaan paperien julkaiseminen lehdessä, jossa David itse toimii päätoimittajana. Näillä toimilla halutaan lisätä DLM-forumin kiinnostavuutta ja saavutettavuutta.



*John Sheridan* UK:n kansallisarkistosta puhui keynote-puheenvuorossaan vahingollisen sähköisen arkiston luomisesta ja ylläpitämisestä. Lähtökohtana puheessa olivat monet ohjeet jopa sadan vuoden takaa: ohjeet siitä kuinka arkistointia pitäisi tehdä ja kuinka nämä paperimaailman toimintamallit ovat luoneet ongelmia sähköiseen arkistointiin. Paperiajan seulonta, kuvailu, säilytys, yms. käytännöt eivät vain toimi nykyisten sähköisten arkistojen kanssa. Jonkin vanhan ohjeistuksen mukaan esimerkiksi sotkuisen verkkolevyn siistiminen arkistointia varten muuttaisi alkuperäistä tietoa ja olisi täten arkistointikäsitteiden vastaista.

Kaikki verkkolevyjä penkoneet tietävät, että kaikkea niillä makaavaa rojua ei todellakaan voi tunkea pitkäaikaissäilytykseen. Toiseksi suureksi ongelmaksi John totesi tiedontuottajien kyvyn säilyttää tietoa edes niin kauan, että sen voisi luovuttaa arkistoon. Säilyttäminen kun harvoin kuuluu tiedontuottajan substanssi-osaamiseen. Tähän ratkaisuna voisi olla sähköisen kopion ottaminen arkistoon heti kun asiakirja/dokumentti yms. on niin sanotussa lopullisessa tilassaan.

## Lohkoketjuista

Seuraavassa kokonaisuudessa käsiteltiin elinkaaren hallintaa sekä sähköistä säilyttämistä hyödyntäen lohkoketjutekniikkaa. Lohkoketjutekniikan perusteiden jälkeen päästiin asiaan. *Lars Hansen* ruotsalaisesta Enigio Time yrityksestä kertoi sähköiseen säilyttämiseen liittyvästä laista, joka on useassa Euroopan maassa teknologianeutraali. Laki siis määrittää vaatimukset, mutta ei mediaa. Pelkästään tällä perusteella lohkoketjukiin on varteen otettava vaihtoehto, jo esimerkiksi sen tarjoaman turvallisuuden ja autenttisuuden vuoksi.

Hansen myöskin totesi, ettei kannata vaatia lainsäätäjiä muuttamaan lakeja uuden teknologian takia, vaan pyrkiä käyttämään teknologiaa siten, ettei lakia tarvitse muuttaa.

AIT:n *Ross Kingin* mukaan lohkoketju tarjoaa erinomaisen pro-venienssin ja pysyvyyden, mutta yksityisten lohkoketjujen tapauksessa hajautus voi jäädä turhan pieneksi. Lopuksi Ross totesi, että lohkoketjutekniikkaa valitessa on syytä olla tarkkana. Jokin firma voi esimerkiksi väittää omaa yksityisen lohkoketjun tarjoamaa tuotettaan super-turvalliseksi, mutta Ross kehoitti kysymään, miten asian voi todentaa, jollei ratkaisua ole testattu maailman mittakaavassa.

Tuttavamme *Kuldar Ass* väitti olevansa hermostunut, koska hän puhui ensimmäistä kertaa moneen vuoteen jostain muusta kuin E-arkista ja toisekseen, koska ei ole lohkoketjuammattilainen. Kuldar kertoi, että Virossa on kokeiltu lohkoketjuja jo kolmen vuotta. E-health puolella kansallinen IT-turvamääräys vaatii hash-ketjuja tiedon eheyden takaamiseen. Tässä on käytetty lohkoketjuja, mutta ongelmaksi muodostuu noin 40 000 uutta dokumenttia ja noin miljoona uutta tapahtumaan päivässä, jotka kaikki pitää todentaa. Kun yhden lohkoketjun laskenta voi viedä pahimmillaan tunteja, on skaalautuvuusongelma valmis.

Virossa ongelma on ratkaistu laskemalla hasheja hasheista ja tekemällä lohkoketjuvarmennukset vain ylimmän tason hasheille. Virossa hyödynnetään valtiojohdon ylläpitämää lohkoketjunodeverkostoa ja simpeliä APIa, jonka käyttäminen on kaikille virallisille toimijoille ilmaista. Lopuksi Kuldar totesi luottamuksen olevan muuttuva käsite: ensin oli allekirjoitus, sitten hash, sitten digi-

taallinen allekirjoitus, nyt lohkoketju ja tulevaisuudessa luultavasti jotain uutta ja parempaa. Arkistoinnissa IP-paketeissa voisi lohkoketjulla korvata normaalin hashin tai digitaalisen allekirjoituksen.

Samaa asiaa käsittelee Zagrebin yliopiston *Hrvoje Stancicin* puhe, joka pohjautui interPARES trust hankkeessa tehtävään tutkimukseen digitaalisen allekirjoituksen tallentamisesta. Ongelma digitaalisten allekirjoitusten kanssa on voimassa-oloaika, joka yleisesti on vain yhdestä kahteen vuoteen. Toinen ongelma on erinäiset käytännöt ja teknologiat allekirjoituksessa. Ehdotettu ratkaisu on tallentaa digitaalisen allekirjoituksen validointitieto lohkoketjuun allekirjoitusvaiheessa, jolloin itse allekirjoitusta ei enää tarvittaisi. Lopuksi Stancic ehdotti, että luotettavat toimijat (lue: arkistot) perustaisivat luottoketjun nodeineen, jotka varmentaisivat toisiaan.

Esitysten jälkeen oli vuorossa lohkoketjupaneeli. Muita järkeviä kysymyksiä ei allekirjoittaneen mielestä esiintynyt kuin ”kuinka oikeus tulla unohdetuksi toteutuu lohkoketjutekniikkaa hyödynnettäessä”. Kuldar totesi yksiselitteisesti, että sellaista materiaalia, jossa voi tulla vastaan oikeus tulla unohdetuksi, ei laiteta lohkoketjuun. Lohkoketjutekniikkaa tunteemattomille mainittakoon, että ketjua muutettaessa kohdasta X, kaikki tuon kohdan jälkeiset pisteet muuttuvat epäosiksi, ellei niitä kaikkia lasketa uudelleen. Käytännössä tämä ketjujen uudelleenlaskenta on kuitenkin mahdotonta.

## Sähköpostien arkistointia

Päivän viimeinen sessio oli otsikoitu uusiksi lähestymistavoiksi tiedonhallintaan. Sisältö koostui kuitenkin vanhoista perusongelmista: tärkeää



tietoa useissa eri järjestelmissä, muutosvastarinta työntekijöiden keskuudessa, IT ja business eivät keskustelee, suurempi tietomäärä tuo suuremmat tietovuodot, tiedon arvo ei ole vakio, automaattisen luokittelun ongelmat jne. Tartun näistä tarkemmin automaattiseen luokitteluun, josta *James Lappin* kertoi. Hän esitteli kehittämäänsä algoritmia, jolla tärkeäksi luokiteltavat sähköpostit saataisiin suoraan asianhallintajärjestelmään. Lähtötilanteessa hänen konsultoimansa firman myyntihenkilöstön oli itse lisättävä liiketoimintaan liittyvät sähköpostit asianhallintajärjestelmään.

Alkutilanteessa työntekijät lähettivät keskimäärin 33 sähköpostia päivässä, joista 0.5 sähköpostia pääsi asianhallintajärjestelmään. Tilannekartoituksen aikana työntekijöiden piti sähköpostin lähetettyään vastata kysymykseen, oliko kyseessä liiketoimintaan liittyvä sähköposti vai ei. Keskimäärin 22 viestiä päivässä liittyi liiketoimintaan. Jotain oli siis tehtävä.

Lappin kertoi algoritminsa tarkkuuden olevan 99,9 % henkilökohtaisten sähköpostien seulonnassa. Businesshenkilöiden mielestä tämä oli loistava saavutus, mutta riskienhallintahenkilöstö näki asian toisin: 200 työntekijää, 30 sähköpostia päivässä -> 6 henkilökohtaista sähköpostia päivässä päätyisi asianhallintajärjestelmään -> yli 1500 henkilökohtaista sähköpostia vuodessa päätyisi asianhallintajärjestelmään.

Seuraavaksi puhe kääntyi poliittisesti arkoihin sähköposteihin ja sitten mielipiteitä sisältäviin sähköposteihin. Koko case oli todennäköisesti fiktiivinen, mutta se kertoo erittäin hyvin näkökulmien erot. Jos on pienikin mahdollisuus virheeseen, mitään ei tapahdu, koska pelätään mahdollisia sanktioita. Tästä ajattelutavasta pitäisi päästä eroon tai junnauksemme kaikki paikoillaan.

## Xamk:n esitykset

Viimeinen mainitsemisen arvoinen puheenvuoro ennen Xamkkilaisten esityksiä oli E-arkin jälkimainingit ja tulevaisuuden näkymät. Puhujat olivat vanhoja tuttuja E-arkin parissa hääränneitä henkilöitä. E-ark ei ole kuulemamme mukaan projekti, joka loppuessaan katoaa, vaan monet E-arkin aikana kehitetyt toiminnot ja palvelut ovat jääneet käyttöön ja tuottaneet monia taloudellisia ja käytännöllisiä hyötyjä. Työkaluja kuitenkin tulee ja menee, ja siksi onkin ensiarvoisen tärkeää säilyttää yhteensopivuus työkalujen välillä. Tämän yhteensopivuuden pohjana ovat määritelmät ja standardit. E-arkissa kehitettyjä määritelmiä ja standardeja onkin asetettu vaalimaan DAS (Digital Archiving Standards) board. Rahoituksen jatkosta on neuvoteltu muun muassa CEF building blocks ohjelman kanssa (tarkemmin sanottuna CEF digital ohjelma). Näillä näkymin sähköinen arkistointi päässee mukaan 01/2018 alkavalle toimintakaudelle, mutta täyttä varmuutta tästä ei vielä ole.

Xamkin delegaatilla oli konferenssissa kaksi esitystä. Ensin kansalaisarkiston ideologisesta näkökulmasta ja MyData-periaatteesta kertoivat Talsi & Lampi. Tämän jälkeen allekirjoittanut kertoi hieman tarkemmin teknisestä taustasta sekä muutamista toiminnoista, joita kansalaisarkistoon on jo integroitu tai tullaan integroimaan. Näistä mainittakoon Failissa aiemminkin esitelty sähköpostikonversio, automaattinen PDF-tiedostojen pilkonta sekä automaattinen OC-tunnistus sellaiselle materiaalille, jossa kyseistä tekstitietoa ei ole.

Loppukaneettina totean, että niinkin ”pieni” tutkimuslaitos kuin CERN käyttää asianhallintajärjestelmään Alfrescon ilmaista community editionia. Eli aina ei siis ole tarpeen maksaa kaupallisista tuotteista vaan ilmaisellakin pärjää, jos on itsellään kykyä ja tahtoa muokata toiminnallisuksia sopivaksi.

## Seuraava DLM Forum

Seuraava DLM Forumin jäsenkokous järjestetään loppukevästä 2018 Virossa. Viimeistään sinne kaikki suomalaisjäsenet mukaan, koska enää ei voi matkakustannuksiakaan pitää esteenä.