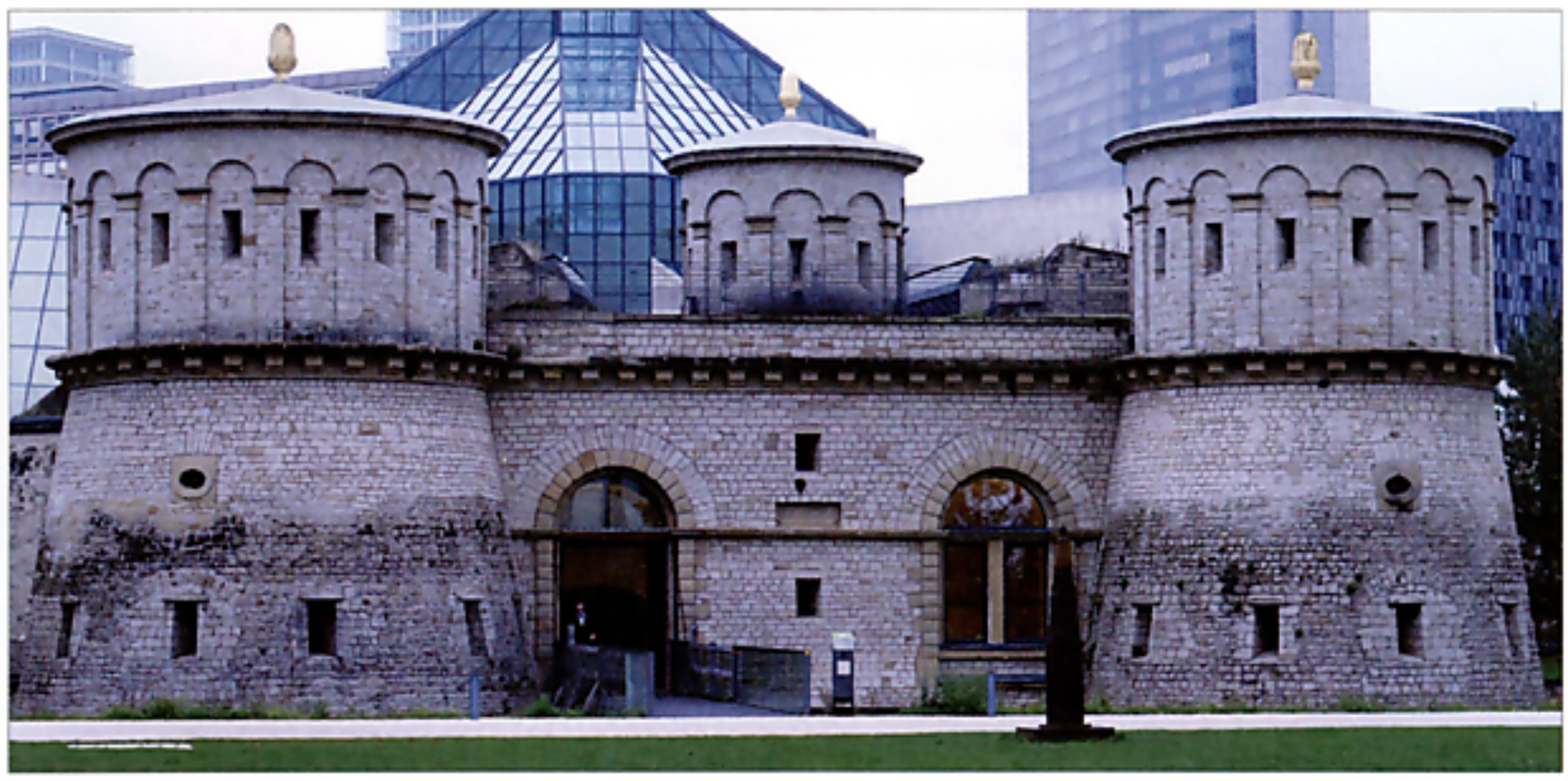
Anssi Jääskeläinen
Yliopettaja
Mikkelin Ammatti-korkeakoulu

DLM Forum kuuli komission terveisiä



Yhden kerran tauon jälkeen palasin DLM-Forumin puolivuosittain pidettävään jäsenkokoukseen. Syynä väliin jäämiseen oli toimiminen opetustehtävissä. Mamkin edustaja oli tässäkin kokouksessa, mutta Failiin asti kokemukset eivät yltäneet.

Kokous pidettiin Musée Dräi Eechelenissä (linnoitusmuseo, kuvassa), josta aukeni hienot näkymät Luxemburgin kaupungin suuntaan 14-15.10.2015. DLM-Forumin puheenjohtaja *Zoltán Szatucsek* avasi kokouksen ja piti perinteisen kiitospuheen järjestäjille sekä esitteli hieman Luxemburgia. Vaikka kyseessä on pieni maa, kansallisarkiston uumenissa on 45 hyllykilometriä aineistoa ja vanhin materiaali on niinkin kaukaa kuin vuodelta 1762. Puheessaan Zoltán mainitsi, että Luxemburg on aina ollut hyvä rakentamaan siltoja, sen vuoksi tämän kertaisen DLM-Forumin pääaihe olikin sillanrakennus (building bridges). Seuraavan puheenvuoron sai Musée Dräi Eechelenin *François Reinert*, Conservateur délégué à la direction, jonka titteliä en edes yritä suomentaa.

Varsinainen sisältö alkoi sessiolla, joka käsitteli sähköistä asiakirjahallintoa EU instituutioissa. Ensimmäiseksi puhui komission *Annemieke Vanlaer*, joka kuulemma tietää kaiken komission sähköiseen säilyttämiseen liittyvistä asioista. Puhe keskittyi suurilta osin heidän Hermes RM järjestelmäänsä ja siihen liittyviin säilytyspalveluihin. Ennen vuotta 2008 Euroopan komissiossa oli käytössä hajautettu asiakirjahallinto, joka yksinkertaisesti tarkoittaa sitä, että jokaisella instituutilla oli omansa. 2008 siirryttiin keskitettyyn malliin ja kaikki asiakirja-aineistot, joilla on oma säilytysaikasuunnitelmansa (retention policy) arkistoidaan Hermeksen avulla. Tämän vuoden aikana Hermekseen on tarkoitus toteuttaa Preservica kustomoinnit ja integraatiot, jonka jälkeen toteutettaneen julkinen pääsy aineistoon. Toteutuksessa on noudatettu OAIS mallia, jolla on varmistettu, että kaikki osapuolet puhuvat ns. samaa kieltä. Hyödyksi työssä on käytetty mm. E-Ark projektin tuottamia raportteja sekä määrityksiä.

Euroopan komission *Yvo Volman* kertoi puheessaan siitä kuinka komissio pyrkii tiedon julkaisijasta informaation keskipisteeksi. Tavoitteen saavuttamiseksi tarvitaan paljon uutta osaamista, jota on kuulemma vaikea löytää koska kaikki parhaat asiantuntijat ovat jo kaupallisten yritysten palkkalistoilla. Hän myös kertoi että europa.eu sivut haravoidaan neljä kertaa vuodessa Internet Memory Foundationin toimesta, jotta myös sivustoilla oleva tieto saadaan arkistoitua. Hän mainitsi myös sen faktan että the Waybackmachine arkistoi sivut myös, mutta kysyi pitäisikö meidän luottaa ainoastaan amerikkalaisen firman toimintaan. Puheessaan hän sivusi myös digitaalisuuden megatrendejä, digitaalista hallintoa, Internetiä, tapaa jolla tietoa kulutetaan sekä avointa tietoa. Hänen esittämiensä laskelmien mukaan avoimen tiedon hyödyt voivat olla vuositasolla jopa 40 miljardia euroa. Luonnollisesti esiin nousi myös laki tiedon uudelleen käytöstä julkisella sektorilla ja esim. US executive order, jonka mukaan hallinnollinen tieto on oletuksena avointa ja koneluettavassa muodossa. – *Koskahan tämä käsittää myös NSA:n tiedot, jotta seuraavan Snowdenin ei tarvitse niitä vuotaa? – Kirjoittajan huomautus.*

Loppukeskusteluissa nousi esille kysymys, miksi päättäjät eivät halua avata tietojaan. Vastauksena tarjottiin tyypilliset kolme syytä: muutosvastarinta; pelko, että joku voi tehdä tiedoilla rahaa (eikä siis

tiedonhaltija itse) ja viimeisenä pelko, että joku tulkitsee tietoja väärin. Yvo kuitenkin kehotti vahvasti siihen, että tiedot avataan ja annetaan ihmisten tehdä niillä mitä he haluavat. Allekirjoittanut on tästä asiasta samaa mieltä, ei ole tiedon haltijan tehtävä päättää mitä tiedolla voi tehdä ja mitä ei, vain avaamalla tiedot ja antamalla ne kansalaisten käyttöön ilman rajoituksia voi syntyä uusia innovatiivisia tapoja hyödyntää ja yhdistellä tietoja.

Euroopan komission osuuden lopetti *Alina Senn* luovuusyksiköstä. Hän kertoi puheessaan mahdollisesti DLM-Forumia ja kuulijoita kiinnostavista Horisontti 2020 rahoituksista. Mielenkiintoisimpina rahoitusmahdollisuuksina nousivat esille yleisesti koko societal challenge 6 work programme 2016-2017 sekä tästä erityisesti Co-creation-2017-4 kutsu. Tästä mainittakoon esimerkiksi rahoitus, jota voi saada virtuaalisiin museoihin liittyviin hankkeisiin. Hän totesi, että Horisontti-rahoituksessa painotetaan enemmän innovaatioita kuin edellisessä seitsemännessä puiteohjelmassa koskaan tehtiin.

"Täydellisen tiedon halu voi helposti johtaa tilanteeseen jossa kaikki mahdollinen arkistoidaan ja käydään 'sitten aikanaan' lävitse."

Kahvitauon jälkeen vuorossa olivat vakiokalustoon kuuluvat *Tim Callister*, Euroopan ARMAN johtaja sekä asiakirjahallinnon konsultti *James Lappin*. Tim puhui kultakutreista (Goldilocks) eli siitä kuinka täydellisen tiedon halu voi vaikuttaa myös hallitustason tiedonhallintaa koskeviin ohjelmiin. Täydellisen tiedon halu voi helposti johtaa hänen mukaansa tilanteeseen jossa kaikki mahdollinen arkistoidaan ja käydään "sitten aikanaan" lävitse. Ideaalimaailmassa ehkä, mutta käytännössä organisaatiodata tulee Timin mukaan kasvamaan noin 650% seuraavan viiden vuoden aikana, joten todellisessa maailmassa kukaan ei koskaan käy arkistoituja tietoja kunnolla läpi.

Tähän liittyen hän antoi esimerkin Hollannista, jossa viralliseen tiedusteluun, olipa se kenen tahansa tekemä, on vastattava tietyn ajan kuluessa. Jos vastausta ei tule annetussa ajassa, kysyjä saa jonkinlaisen kompensaation. Hollannissa on kuulemma henkilöitä jotka elättävät itsensä näillä kompensaatioilla. Siksi hyvä tiedonhallinta, löydettävyys ja jaettavuus ovatkin ensiarvoisen tärkeitä asioita. Tietojen avaamisesta Timillä oli myös mielipiteensä, hän oli jaotellut syyt jakamattomuudelle muutamien elokuvista tuttujen repliikkien mukaan:

- **Käyttäjänäkökulma:** *"This is my information. There are many like it, but this one is mine."* Vapaasti suomennettuna: Tämä on minun tietoni, samanlaisia on monia mutta tämä on minun. (Lausahdus tunnetaan parhaiten elokuvasta Full Metal Jacket, jossa Vincent D'Onofrion esittämä sotamies Leonard 'Gomer Pyle' Lawrence sekoaa).
- **Lakinäkökulma:** *"My preciousssssssssssssssss"*. Vapaasti suomennettuna: Minun kallisarvoisinnnnnnnn. (Tunnetaan elokuvasarjasta Taru sormusten herrasta, jossa klonkku viittaa tällä sanonnalla mahtisormukseen).
- **IT näkökulma:** *"The servers cannae take anymore captain"*. Vapaasti suomennettuna: Palvelimet eivät kestä enempää kapu. (Todennäköisesti elokuvasta Star Trek, jossa Scotty viittaa aluksen kestävyyteen taistelussa)

Myös Tim kehotti, että tiedon annetaan kertoa mitä se haluaa kertoa. Myös Bigdata sai arvostelua osakseen: hän kysyi mikä on Bigdataa? Kuulaskeutumisdata on määrällisesti erittäin pieni mutta ihmiskunnan historian kannalta ehkä yksi tärkeimmistä, onko se Bigdataa? Lopuksi hän tavastaan poiketen esitteli muutamia avoimesti saatavilla olevia työkaluja digitaalisen tiedon riskien kartoitukseen. Näistä mainittakoon nyt Preservican Digital value at risk calculator.

Lappinin puheenvuoro oli jälleen höystetty käsin piirretyillä kuvilla, tosin tällä kertaa lienee kiire iskenyt esitystä tehdessä, koska piirroksia oli mielestäni vähemmän kuin aiemmin. Puhe käsitteli sähköpostien oikeaoppista arkistointia, koska niiden määrä lisääntyy jatkuvasti ja arvosisältö vähenee. Varsinaista juuta tai eitä ei esityksessä löytynyt, mutta pääasiallinen sanoma oli että pidetään kaikki ja hoidetaan seulonta sitten joskus aikanaan autoclassification työkalujen avulla.

Torstain ensimmäisessä sessiossa oli mukana DLM-Forumin uusi jäsen Novo Nordisk:in edustaja *Torben Juul*. Hän kertoi lääkefirman näkökulmasta erilaisten maakohtaisten lakien, normien, määräysten yms. vaikutuksesta lääkefirman toimintaan.

Toisena oli odotettu Jon Garde ja Moreq2010 katsaus, mutta Jon oli joutunut jäämään pois kokouksesta henkilökohtaisten syiden takia. *José Borbinha*, Portugalin Instituto Superior Técnicosta paikkasi tilannetta parhaansa mukaan. Moreq2010 suhteen on menossa kaksi rinnakkaista kehitystyötä. Jon

kehittää Moreqia ja sen testiskriptejä, joiden pitäisi olla valmiit vuoden 2015 aikana. Hän myös kertoi että Moreq2010 moduleitten draft-versioiden arvioijaksi voi ilmoittautua.

Moreq2010 ympärille ollaan saman aikaisesti rakentamassa Moreq online -palvelua. Jo olemassa oleva Moreq2010 vientityökalun validointiympäristö linkitetään suoraan tähän Moreq online -palveluun. Jos käyttäjä siis haluaa testata MYJ:n (Moreq Yhteensopiva Järjestelmä) vientitoimintoa hän voi validoida vientityökalun tuloksen validaattorilla ja nähdä suoraan Moreq online palvelussa läpimenneet ja epäonnistuneet validoinnit.

Tietosuojadirektiivin nojalla lähes kaikki toiminta mitä kirjastoissa ja arkistoissa tehdään, on laitonta.

Torstaiaamupäivän viimeinen esitys oli Portsmouthin yliopiston digitaalisen humanismin professorin *David Andersonin*. Hän kertoili EU:n uudesta tietosuojadirektiivistä ilman lakimiestaustaa. Hän mm. mainitsi että kyseisen lain nojalla lähes kaikki toiminta mitä kirjastoissa ja arkistoissa tehdään, on laitonta. Lisäksi olemassa oleva EU lainsäädäntö toisaalta pakottaa tietojen uudelleen käyttöön, mutta tämä tietosuojadirektiivi puolestaan lyö kapuloita rattaisiin. Lisäksi esille nousi oikeus tulla unohdetuksi, joka tulee koskemaan erityisesti hakukoneita. Toisaalta allekirjoittanut on vahvasti sitä mieltä, että Google ja muut suuret amerikkalaiset toimijat tekevät edelleen mitä haluavat riippumatta EU:n direktiiveistä. Nähtäväksi jää osuuko veikkaukseni oikeaan vai ei.

Kahvitauon jälkeen oli vuorossa ehkä odotetuin sessio, joka käsitteli E-ark projektia. Session avasi projektin koordinaattori *Janet Delve*. Lyhykäisyydessään kaikki (aikataulut, budjetti, projektin tuotokset, viestintä, yms.) menee kuten pitääkin.

Janetin alustuksen jälkeen *Kuldar Aas* Virosta kertoi tarkemmin asiasta nimeltä common specification, jotka ovat yleismäärityksset jokaiselle E-arkin IP paketille (luovutuspaketti SIP, arkistopaketti AIP, ja jakelupaketti DIP). Jokaisen paketin tulee noudattaa yleismäärityksen mukaista fyysistä rakennetta ja toteuttaa sen vaatimukset. E-arkin julkaistuissa tuotoksissa löytyy myös tästä asiasta lisätietotoja. Kuldarin jälkeen DLM-Forumin varapuheenjohtaja *Ricardo Vieira* kertoi kuulijoille asiakirjahallinnon kypsyysmallista, jonka suhde E-arkiin jäi ainakin allekirjoittaneelle hieman epäselväksi.

Viimeisenä ja kaikkein teknisempänä puheenvuorona sessiossa oli Austrian Institute of Technologyn *Sven Schlarbin* puheenvuoro E-arkin AIP paketeista ja E-arkiin integroitavasta referenssitoteutuksesta joka tulee perustumaan Hadooppiin, Luceneen, SolRiin, Tikaan, Lilyyn sekä Hbaseen. Pääasia E-arkin AIP paketissa on se, että alkuperäinen SIP paketti sisällytetään AIP:iin sellaisenaan kuin se otettiin vastaan, näinollen vertailu missä tahansa säilytyksen vaiheessa on mahdollinen.

Avoimen lähdekoodin käyttäminen referenssitoteutuksessa sekä yhteiseen AIP rakenteeseen ohjaaminen herätti etenkin kaupallisen toimijan edustajassa vahvoja mielipiteitä. Hän oli esim. sitä mieltä että AIP paketin tulisi olla järjestelmäkohtainen ja mahdollisimman tehokkaasti toteutettu. Sven totesikin, että E-arkin tarkoituksena ei ole kehittää kilpailevaa avoimen lähdekoodin arkistoratkaisua, ainoastaan referenssitoteutus. AIP paketeista Sven totesi että E-ark ei missään tapauksessa oleta kaikkien toteutuksien AIP pakettien olevan täysin yhteensopivia. E-ark projektin tapauksessa riittää että AIP paketit noudattavat yleismäärityksiä. Svenin mukaan tärkeintä on, että DIP paketit ovat täysin määrittelyjen mukaisia ja siksi siirrettävissä helposti järjestelmien välillä. Tämä on niin kutsuttu exit strategy, joka myös suomalaisten toimijoiden tulisi pitää mielessä tietojärjestelmiä valittaessa. (*Exit strategy eli aineiston siirtostrategia järjestelmän käytön loppuessa tarvitaan nimenomaan arkistopaketeille – Päätoimittajan huomautus.*)

Pääasia E-arkin AIP paketissa on se, että alkuperäinen SIP paketti sisällytetään AIP:iin sellaisenaan kuin se otettiin vastaan, näinollen vertailu missä tahansa säilytyksen vaiheessa on mahdollinen.

Päivän viimeisessä sessiossa oli kaksi esitystä. Ensimmäisen piti Sveitsin Bundesarchivin *Krystyna Ohnesorge*. Hän puhui tietokantojen arkistointiin käytettävän SIARD formaatin versiosta 2.0 ja sen mukanaan tuomista eduista. Taustalla oli perusongelma siitä että tietokantojen suljettuja formaatteja ei voida arkistoida sellaisenaan koska on hyvin epätodennäköistä että niitä ymmärtäviä ohjelmia löytyy vielä 10-20 vuoden päästä.

Ratkaisuna on normalisointi SIARDia käyttämällä. Vanha SIARD perustuu kuitenkin sql:1999 määrityksiin joten päivitystä kaivattiin. Uudistettu versio perustuu sql:2008

määrityksiin ja tuo mukanaan mm. mahdollisuuden tietokantojen pakkaamiseen, tuen käyttäjän määrittelemille tietotyypeille sekä tuen BLOBeille (Binary Large Object) and CLOBeille (Character Large Object). Mikä parasta SIARD2 on taaksepäin yhteensopivat SIARD1:n kanssa joten tarvetta migraatiolle ei ole.

Viimeisenä naapurimaamme Ruotsin kansallisarkiston teknineen neuvonantaja *Karin Bredenberg* kertoi siitä kuinka hallituksen e-arkistointi hoidetaan Ruotsissa. Pääasia puheessa oli kaikkien erilaisten lakien, normien ja määräyksien huomioon ottaminen ja ehkä rivien välistä oli luettavissa toive näiden järkeistämisestä.

Tämänkertaisesta DLM:n tapaamisesta jäi hieman ristiriitainen tunne. Sinänsä oli mukava taas tavata vanhoja tuttuja, mutta jollakin tavalla tunnelma on myös muuttunut. Liekö Garden poissaolo ja Moreq2010:n mielestäni lähes muuttumaton tilanne sitten 2014 vuoden viimeisen kokouksen syinä tähän.

Seuraava DLM-Forumin tapaaminen pidetään Haagissa Hollannissa kesäkuun 8-9 päivä. Täällä yleisinä aiheina lienevät tosielämän standardit, innovaatiot, kiinnostavat projektit sekä koulutus. Jos aiheet pysyvät näinä, voi hyvinkin olla että Mamkista saadaan taas vaihteeksi puheenvuoro DLM-Forumiin. Allekirjoittanut jää odottelemaan kokousta sekä vuoden loppuun luvattuja Moreq2010 testiskriptejä. F