



Kaakkois-Suomen
ammattikorkeakoulu



South-Eastern Finland
University of Applied Sciences

**PLEASE NOTE! THIS IS A PARALLEL PUBLISHED VERSION /
SELF-ARCHIVED VERSION 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-Forum Ljubljanaassa

Version: Publisher's PDF

Please cite the original version:

Jääskeläinen, A. (2023). DLM-Forum Ljubljanaassa. Faili 3, 24 - 27.

HUOM! TÄMÄ ON RINNAKKAISTALLENNE

Rinnakkaistallennettu versio voi erota alkuperäisestä julkaistusta sivunumeroiltaan ja ilmeeltään.

Tekijä(t): Jääskeläinen, Anssi

Otsikko: DLM-Forum Ljubljanaassa

Versio: Publisher's PDF

Käytä viittauksessa alkuperäistä lähdettä:

Jääskeläinen, A. (2023). DLM-Forum Ljubljanaassa. Faili 3, 24 - 27.

DLM-Forum Ljubljanassa



Anssi Jääskeläinen
Tutkimuspäällikkö
Xamk
Digitalia

Suomesta oli DLM Forumin Ljubljanan tapaamisessa tällä kertaa mukana peräti neljä matkajaa, kaksi Xamkista, yksi Elkasta ja yksi Kansallisarkistolta. Tämä oli uuden jäsenen Elkan ensimmäinen vierailu DLM Forumissa, tervetuloa mukaan!

Tälläkin kertaa DLM-Forumissa yhteydessä pidettiin eArchiving Geoconference, mutta nyt aikataulullisista syistä tämä tapahtuma jäi allekirjoittaneelta väliin. Ljubljanassa saavuttiin tiistaina 9.5. ja varsinaisen

nen tapahtuma alkoi keskiviikkona 10.5. Tiistain vapaa-aika oli hyvä käyttää tutustumalla yhteen Euroopan suurimmista ostoskeskuksista, BTC City Ljubljanan. Sieltä matka jatkui apostolinkyydillä Ljubljanan linnalle, josta aukesi hyvä näkymät kaupungin suuntaan, ja olihan siellä toki virvokkeitakin tarjolla matkajia varten.

DLM-Forumissa tapaaminen pidettiin Domus Medica -konferenssikeskuksessa. Puitteet tilaisuudelle olivat mainiot – ainoana pienenä miinuksena 3 km etäisyys keskustasta. Keskiviikon aamupäivä alkoi DLM forumin johtokunnan kokouksella, josta ainoana isompaa uutisena on Norjan Kansallisarkiston Espen Sjøvollin eroilmoitus. Hän vaihtaa työpaikkaa ja ei siksi ainoakaan toistaiseksi voi enää toimia johtokunnassa. Lisäksi keskusteltiin Euroopan komission logon

näkyemisestä DILCIS (Digital Information LifeCycle Interoperability Board) ylläpitämässä IP (Information Package) määrittelydokumenteissa, siten määrittelyistä saataisiin ”virallisemmat”, koska kyseessä eivät kuitenkaan ole viralliset standardit. Allekirjoittaneen henkilökohtainen mielipide on komission pitäminen erossa näistä dokumenteista, koska heidän mukanaan olo toisi eteen samat ongelmat, joihin EARK-CSP toiminnassa on törmätty. Eli mitään ei tapahdu ennen varmistuksen varmistuksen varmistuksia monelta eri taholta.

Tapaamisen avasi DLM Forumissa puheenjohtaja Anja Paulic Slovenian Kansallisarkistolta yhdessä Nova Gorican paikallisarkiston Tanja Martelancin kanssa. Tanjan edustama paikallisarkisto on Elkan tavoin DLM Forumissa uusia jäseniä, mutta he olivat kuitenkin jo järjestämässä



Ljubljanan linnasta aukesi hienot näkymät kaupungin suuntaan. Kuva: Anssi Jääskeläinen.



Suomen delegaatio DLM-Forumissa Ljubljanassa: Vasemmalta Anssi Jääskeläinen (Xamk), Markus Meremies (Kansallisarkisto), Kati Saltiola (Xamk) ja Mika Kokkonen (ELKA).

tilaisuutta yhdessä Kansallisarkiston kanssa. Sekä Tanja että Anja toivoivat läsnäolijoilta aktiivista osallistumista ja tiedon jakamista.

Ketterät arkistot?

Ensimmäisen session isännöi John Sheridan (The National Archives UK). Puheenvuoroissa käsiteltiin muun muassa Norjan Kansallisarkiston Espanin toimesta kokemuksia ketterästä arkistosta. Termi herättää varmasti kiinnostusta lukijoissa, koska arkistoja ei perinteisesti mielletä kovinkaan ketteriksi toimijoiksi. Kuten kaikki arkistotoimijat tietävät, tietoja on hankala löytää ja käyttää, kos-

ka kaikki prosessit on suunniteltu paperiajan mallilla. Vertauksena Espen käytti vesiputousmallia, jossa tietää mitä saa, mutta jo valmistuessaan vanhentuneena. Vastakohtana ja kehityksen lähtökohtana olivat ketterät kehitysmenettelyt, mutta toiminto kerrallaan. Puheessaan hän käytti vertausta ”moottoritie tai kaunis vuoristotie”. Jos edetään kaasu pohjassa maaliin asti, sinne päästään nopeasti, mutta lopputulos ei välttämättä miellytä käyttäjiä. Norjalaisten ratkaisussa edetään vuoristotietä ja katsellaan tarkasti, mitä kokemuksia muutoksista saadaan ja muutetaan prosessia tarpeen vaatiessa.

Palataksemme termiin ketterät arkistot, suurimmaksi esteeksi Espen mainitsi asennemuutokset. Arkistoväelle on ollut perinteisesti vaikeaa, ellei mahdotonta laittaa jotakin keskeneräistä kritiikin kohteeksi, ja juuri näin ketterissä kehitysmenettelyissä tulisi toimia. Lisäksi olisi hyväksyttävä se fakta, että johtoportaalte tai asiakkaille ei voida luvata tarkasti, milloin jokin ominaisuus on valmiina, vaan kehitettävät toiminnallisuudet valitaan jokaiselle sprintille erikseen priorisoimalla ne isommasta backlog-listasta. Eli lyhyesti kuvattuna ketterät arkistot tarkoittavat arkistotoimintojen kehittämistä ketteriä sovelluskehitysmenetelmiä hyödyntäen, samaa

mitä esimerkiksi Disec on tehnyt Yksä-arkistopalvelunsa kanssa jo pitkään.

Session toinen puheenvuoro käsiteli digitointiprosessia toimittajan näkökulmasta ja sen pitivät Rok Bojanc ja Igor Zorko, ZZI nimisestä slovenialaisesta firmasta. Pääasias-
sa puhe oli perushuttua laadusta, säilyvyydestä ja autenttisuudesta. Mieleen painettavat asiat esitykses-
sä olivat: digitoinnissa pääkysymys ei ole digitoidaanko, vaan koska ja miten digitoidaan, sekä sähköinen arkistointi, johon digitoinnilla yleensä tähdätään, mikä ei saa vaikuttaa millään tavalla häiritsevästi liiketoimintaprosesseihin.

Seuraava sessio oli jaettu kahteen eri tilaan aihealueiden perusteella. Itse olin kuulolla Merenmiehen moderoimassa osiossa jossa muun muassa Martin Rechterik Tshekin Kansallisarkistolta puhui Twitter (nykyinen X) aineiston SIP-pake-
toinnista. Heidän ratkaisunsa pohjautui Digitalian Tuomo Räisäsen kehittämään työkaluun, jonka toiminnallisuuksia he olivat kehittäneen eteenpäin. Heidän Linux, Postgresql, Apache ja DBView -ratkaisussaan Twitterin kehittäjätilin rajoituk-
sia oli jollakin tavalla kierretty. Jos tämä olisi tehty yrityksenä heillä olisi todennäköisesti Muskin laki-
henkilöarmeija kimpussa, mutta kansallisarkistoilla taitaa olla tiettyjä ”kulttuurihistoriallisen” aineiston keräämiseen liittyviä etuoikeuksia. Toki Martin mainitsi, että tietoja ei näytetä julkisesti vaan ne ovat saatavilla vain tutkijasaleissa – joitakin Twitterin käyttöehtoja tässä kuitenkin poljettaneen.

Seuraavana Vincent Hooft ja Anna van Lankvelt Hollannin Kansallisarkistolta jatkoivat siitä, mihin jäivät

muutama vuosi sitten Tallinnassa. Perusongelmana on se, että Hollannin laki käskää siirtämään tiedon Kansallisarkistolle, mutta jotkin organisaatio tarvitsevat tietoja omissa aktiivijärjestelmissään huomattavasti pidempään, jotkut jopa pysyvästi. Kertauksena: Viron DLM Forumia käsittelevässä jutussa kerrottiin Hollannin maanmittauslaitoksesta (Land registry), jonka toiminta on täysin ristiriidassa lainsäädännön ja Kansallisarkiston vaateiden kanssa. Ratkaisuksi on kehitelty mallia, jossa tiedon tuottajasta tehdään arkisto. Tällaisen statuksen tiedon tuottajalle voi myöntää vain state archivist, tietyillä tiukoilla kriteereillä. Maanmittauslaitoksen pilotti on valmistunut Tallinnan DLM:n jälkeen ja uuden pilotin kokemuksiin perustuvan lain piti olla jo valmis, mutta tavoite on siirretty vuoteen 2025. Prosessia ei haluta pakottaa läpi, vaan siitä testataan mahdollisia ongelma-
kohtia yhdessä hyödyntäjien kanssa. Lopputulemana on tarkoitus olla laki, joka huomio kaikkien osapuolten prosessit ja tietomallit. Session päättivät Willeke de Groot ja Mirjam Bakker (NoordHollands Archief), myös Hollannista. He kertoivat pre-ingest -työkälystä, jolla arkistoon vietävien aineistojen tarkistuksia tehdään. Työkalu on kuulemamme mukaan helpottanut ja nopeuttanut työtä huomattavasti, ja se on saatavilla osoitteessa <https://github.com/noord-hollandsarchief>.

Ennen AGM:ää (Annual General Meeting) olleessa sessiossa keski-
tyttiin pääasiassa terveystietoon, sen pseudonymisoituu hyödyntä-
miseen ja jakamiseen sekä väestön ikääntymisen mukanaan tuomiin haasteisiin. Tärkeimpänä kysymyk-
senä itselleni heräsi: miten pseudonymisoida sellaista terveystietoa, jossa kenttien nimet eivät välttä-

mättä sisällä sitä mitä kentän nimen perusteella voisi olettaa? Esitykses-
sä keskityttiin ainoastaan ideaaliti-
lanteeseen, jossa kentät sisältävät vain sen tiedon, johon ne on tarkoi-
tettu, mutta tähän ei useinkaan
pidä paikkansa.

Päivän päätti AGM, jossa äänestet-
tiin DLM:n johtokunnan uusista
jäsenistä. Ehdokkaita oli peräti
yhdeksän ja paikkoja jaossa kuusi.
Jo johtokunnassa toimivat Elena
Cortés Ruiz, David Anderson, Anja
Paulic ja John Sheridan valittiin
jatkokaudelle. Uusista ehdokkaista
valituiksi tulivat Stefan Kwasnitza
(Swiss Federal Archives) ja Sofia
Särduquist (Swedish National Archi-
ves).

Torstain parasta antia oli AI ”työ-
paja”, jossa Ljubljanan yliopiston
Janez Demšar ja Blaž Zupan tekivät
muun muassa erilaisia visualisoin-
teja datajoukolle käyttäen avoimen
lähdekoodin Orange Data Mining
-ohjelmistoa (<https://orangedatamining.com/>). Sillä voi muun muassa
tehdä luokitteluja teksteille ja ryh-
mitellä niiden sanoja semanttisesti.
Esimerkissä sänky, lakana, tyyny ja
peitto kuuluivat myös kyseisen oh-
jelmiston mielestä samaan ryhmään.
Toimintalogiikka perustuu siihen,
että sanat muutetaan numeroiksi
ja etäisyysiksi toisistaan sopivaa
kielimallia hyödyntäen. Samanlai-
nen toiminta on mahdollista myös
valokuville, mutta pohjalle tarvitaan
hyvä malli ja paljon laskentateho-
ja. Esityksessä myös todettiin, että
isommat kaupalliset toimijat ja
heidän mallinsa luonnollisesti ohit-
tavat ainakin toistaiseksi avoimet
mallit suorituskyvyssä ja tunnistus-
tarkkuudessa. Ohjelmassa on myös
mahdollisuus putkittaa (pipelining)
eri malleja ja toimintoja toisiinsa ja
siten saavuttaa parempia tuloksia.

Digitalia tulee ainakin tutkimaan kyseisen ohjelmiston hyödyntämismahdollisuutta – sen voisi yhdistää vaikkapa HuggingFace-palvelusta löytyvän TurkuNLP:n suomenkielisen GPT3-mallin kanssa ja katsoa mitä tapahtuu.

Työpajaa seurasi luonnollisena jatkumona AI-aiheinen paneelikeskustelu, jossa myös allekirjoittanut oli mukana. Paneelin kysymykset oli tosin mielestäni suunniteltu negatiivisesta näkökulmasta ja tekoäly nähtiin jonain sellaisena, jonka kanssa on opittava elämään. Kes-kustelussa olivat lisäksi mukana Espen moderaattorina, Samo Zorc (Ministry of Digital Transformation, Slovenia), John Sheridan (The National Archives) ja jo työpajan pitänyt Blaž Zupan. Tästä osiosta saa parhaiten kuvan katsomalla talenteet DLM-Forum sivuilta, tosin ne ovat saatavilla vain jäsenille.

Preservican puheenvuorossa Peter Anderton totesi jälleen, ettei ole myymässä mitään, mutta hän totesi kyseisen ohjelmiston auto preservation -toiminnon olevan helppo tapa ylläpitää isoa tiedostojoukkoa ilman isoa tiimiä. Hän myös totesi Preservicassa olevan nyt tallessa noin 100 miljoonaa tiedostoa, joista noin 75% on maailmalla eniten käytettyjä tiedostomuotoja. Lopussa 25% on kuulemamme mukaan noin 1200 eri formaattia, joille kaikille on olemassa migraatiopolut.

Loppuohjelmassa oli erilaisia lightning talkkeja. Näissä ei mielestäni ollut mitään Failin lukijoita kiinnostavia asioita – todettakoon vain se, että hekin olivat huomanneet harvinaisempien tiedostoformaattien arkistoinnin olevan ongelmallisempaa kuin usein käytettyjen tiedostomuotojen.

Viimeisessä sessiossa ennen Salamancan tapahtuman esittelyä Sari Järn ja Kati Saltiola kertoivat Memory Campuksesta, Memory Labista sekä Digtointin osaajakoulutuksesta ja siihen liittyneestä osaamiskartoituksesta. Näistä kiinnostuneet voivat olla yhteydessä Sariin tai Katiin. Seuraava DLM-Forum järjestetään Salamancassa Espanjassa 23.–24. lokakuuta ja sen

yhteydessä pidetään Euroopan Unionin DGCNETin kanssa Archiving Initiative summit 25.10. EARK-CSP hankepäiväksi on päätetty 26.10. Lokakuussa olisi siis hieman vajaan viikon reissu tarjolla Espanjaan. Konferenssiin voi edelleen lähettää puheenvuoroehdotuksia täällä: <http://www.dlmforum.eu/index.php/all-events/145-2023-triennial-salamanca>



Kati Saltiola esittelemässä Memory Campusta, David Anderson valvoo ajankäyttöä. Kuva: Anssi Jääskeläinen