



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 Bernissä

Version: Publisher's PDF

Please cite the original version:

Jääskeläinen, A. (2019). DLM-Forum Bernissä. Faili 3, 24 - 27.

HUOM! TÄMÄ ON RINNAKKAISTALLENNE

Rinnakkaistallennettu versio voi erota alkuperäisestä julkaisusta sivunumeroiltaan ja ilmeeltään.

Tekijä(t): Jääskeläinen, Anssi

Otsikko: DLM-Forum Bernissä

Versio: Publisher's PDF

Käytä viittauksessa alkuperäistä lähdettä:

Jääskeläinen, A. (2019). DLM-Forum Bernissä. Faili 3, 24 - 27.

DLM-Forum Bernissä



Anssi
Jääskeläinen
TKI-asiantuntija
Xamk
Digitalia

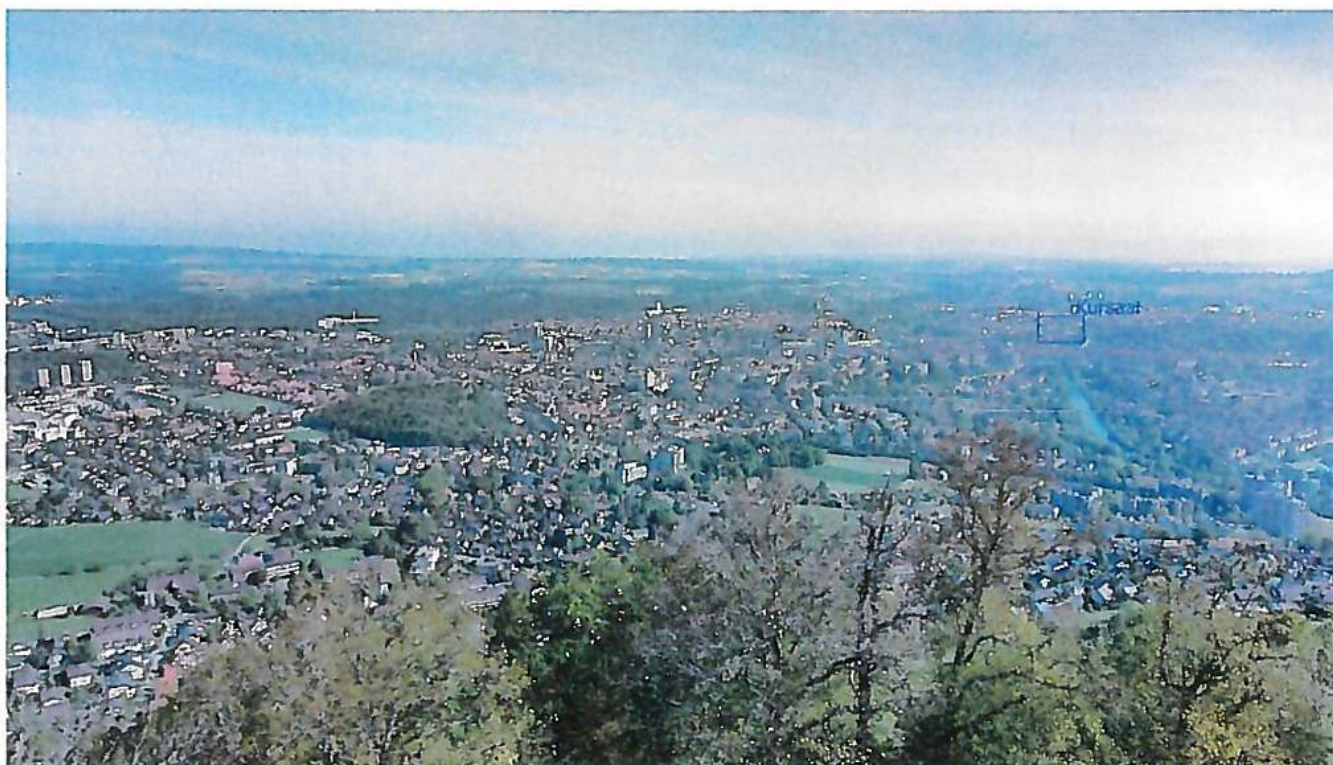
Tällä kertaa DLM-Forummin puolivuositaisen jäsenkokouksen organisoivat Sveitsin kansallisarkisto (Swiss Federal Archives) ja pitopaikkana toimi kongressikeskus/hotelli Kursaal Bern. Puitteet olivat mainiot ja näkymät aukesivat koko Bernin ylle.

Vielä hienomman näkymän kaupunkiin sai kuitenkin kävelemällä Gurten-vuorelle, josta oheinen kuva on napattu. Molempina kokouspäivinä paikalla oli noin 70 kiinnostunutta osallistujaa. Kävijämäärät ovat selkeästi kasvussa, kuten myös tilanne, johon palaan myöhemmin tässä jutussani.

DLM-Forummin puheenjohtaja Jan Dalsten Sørensen avasi jäsenkokouksen ja kiitteli Sveitsin kansallisarkistoa järjestelyistä. Tämän jälkeen virallisen tervetuloiltoilun antoi Philippe Künzler, Sveitsin kansallisarkiston johtaja. Tavallisuudesta poiketen hän oli myös läsnä koko kahden päivän ajan, mikä osaltaan kertoo sekä sisällön kiinnostavuudesta että järjestävän tahon joh-

tajan sitoutuneisuudesta asiaan ja DLM-Forummin toimintaan.

Päivän alustuspuheenvuoron piti Fribourghin yliopiston Philippe Cudré-Mauroux. Hän puhui big datasta ja tietograafien (knowledge graphs) eduista sen hallinnassa. Taustalla oli W3C:n standardoima RDF (Resource Description Framework) ja linkitetty avoin tieto, joiden avulla hänen mielestään saadaan monia etuja perinteisiin hakukoneisiin ja tiedonhakumenetelmiin nähden. Hakuihin saadaan muun muassa älykkyyttä semanttisten yhteyksien avulla. Siitä mitä toteutus varsinaisesti vaatii ei ollut puhetta, mutta heidän ratkaisunsa on yliopistomaisesti vielä kehitteillä.



Bernin kaupunki Gurten-vuorelta (858 m) kuvattuna sekä konferenssipaiikka. Kuva: Anssi Jääskeläinen.

Päivän ensimmäisessä varsinaisessa puheenvuorossa Espen Sjøvoll Norjan kansallisarkistosta puhui siitä kuinka tärkeää olisi huomioida asianmukainen dokumentointi jo suunnitteluvaiheessa. Tietyissä asioissa tällainen lähestymistapa voi toimia, mutta esimerkiksi Digitalia on omassa kehityksessään noudattanut ketterän kehityksen periaatetta ”toimiva softa ennen kattavaa dokumentointia”. Tämä ei tietenkään tarkoita sitä, etteikö dokumentaatiota tarvita, mutta tehtävän dokumentoinnin tulee tukea varsinaista sovelluskehitystyötä eikä hidastaa kehitystiimin työskentelyä.

Seuraavaksi oli Digitalian ”kevennyspuheenvuoro” jossa esittelimme VR (Virtual Reality) -toteutuksen toisen maailmansodan aikaisista venäläisten toteuttamista pommituksista Helsinkiin. Löyhä linkki kokouksen aihepiiriin oli uudenlainen ja kiinnostava esitystapa arkistossa olevien tietojen esittämiseen. Ratkaisu herätti kiinnostusta ja vilkasta keskustelua siitä, mitä arkistoaineistoilla olisikaan mahdollista tehdä. Myös demolaitteisto oli mukana, joten halukkaat saivat itse kokeilla, miltä pommitukset olisivat saattaneet tuntua. Demoa pääsee myös kokeilemaan Digitalian työhuoneella Xamkin Mikkelin kampuksen Mikpoli-rakennuksessa. Verkossa toimiva WebGL-ratkaisu on parasta aikaa toteutuksessa, ja sen valmistuttua sitä pääsee kokeilemaan osoitteessa digitalia.xamk.fi. Suomen Kansallisarkiston Merenmiehen kanssa oli alustavasti puhetta päivitetystä 3D-VR demosta myös syksyn DLM-Forumissa Helsingissä. Jos tämä demo toteutuu, siinä pääsee kokemaan pommitukset Kansallisarkiston ja tuomiokirkon maisemissa.

Päivän viimeisessä puheenvuorossa Sébastien Soyex ja Friedel Geeraert Belgian maakunta-arkistosta (Belgian State Archives) kertoivat heidän Heritrixiin perustuvasta www-arkistoinnistaan. Päivän päätti jäsenkokous, jossa oli monia asioita listalla. Sørensenin ensimmäisenä pitämän toimintakatsauksen jälkeen oli uuden johtokunnan jäsenen valinta. Paikkaan oli tullut kaksi hakemusta, jotka oli lähetetty jäsenistölle etukäteen tutustuttaviksi. Ehdokkaina olivat Markus Merenmies Kansallisarkistosta sekä Espen Sjøvoll Norjan kansallisarkistosta. Molemmat saivat ennen äänestystä pitää vaalipuheensa, jossa molemmat enempi vähempi kannattivat kansainvälisen yhteistyön lisäämistä. Äänestyksen jälkeen valituksi tuli Markus, joten DLM:n johtajistossa on nyt mukana myös suomalaisväriä. Onnittelut Markukselle!

Seuraava suuri muutos oli virallisen nimen muuttaminen vastaamaan paremmin nykyaikaa. Vanhasta nimestä, Document lifecycle management, siirryttiin nimeen Data Lifecycle Management. Perusteluja ei tarvinnut sen enempää, kuin että tieto on nykyisin muualla kuin yhdessä dokumentissa ja dokumentin käsite on hämärtynyt. Lisäksi äänestettiin muutamista sääntömuutoksista, jotka tulevat vaikuttamaan muun muassa jäsenien äänimäärän ja jäsenmaksun suhteeseen. Kaikki muutokset hyväksyttiin yksimielisesti. Lisäksi keskusteltiin ja äänestettiin sääntömuutoksesta, joka olisi tuonut kaikki johtokunnan asiakirjat ja niiden liitteet jäsenille täysin julkiseksi ilman rajoitteita. Tämä ehdotus sai kannatusta, mutta palautettiin kuitenkin käsittelyyn suurelta osin GDPR:n asettamien rajoitteiden vuoksi.

DLM-Forumien taloustilannetta esitteli yleensä rahastonhoitaja David Anderson. Tilanteesta, jossa muutama vuosi sitten oltiin vielä konkurssin partaalla, on noustu tilanteeseen, jossa tileillä on tällä hetkellä reilut 100 000 €. Siirtymä Viroon on edelleen kesken, joten talousraportti oli suhteellisen monimutkainen, mutta tähän on kuulemamme mukaan tulossa muutos ensi vuonna, jolloin pankkisiirtymät on saatettu onnellisesti loppuun. Jäsenistöltä kyseltiin hyvistä ehdotuksista, miten tätä kertynyttä pääomaa voitaisiin käyttää DLM-Forumien toiminnan kehittämiseen ja yhteistyön syventämiseen erinäisten tahojen välillä.

Gaalailallinen oli jälleen loistava niin tarjoilun kuin tunnelmankin puolesta. Xamkin delegaatio – minä ja kollega Teemu Hänninen – juurtui Latvian kansallisarkiston edustajien (arkiston varajohtaja Gatis Karlsons & arkistohallintayksikön päällikkö Liga Vibure) kanssa samaan pöytään. Muiden keskustelujen lomassa keskusteltiin myös mahdollisesta yhteistyöstä liittyen 3D-objektien esittämiseen pohjautuen Digitalian VR-esitykseen. Digitalian seuraavaan hankkeeseen on suunniteltu tähän liittyen 3D-skannerin hankintaa ja virtuaalimuseokonseptin toteuttamista.

Keskiviikon aamupäiväsessiossa oli uusien jäsenien esittelyt ja sisältö. Ivar Rummelhoff Norjalaisesta Piql-yhtiöstä puhui kuolemattomasta virtuaalikonkehuksesta (immortal Virtual Machine, iVM), joka on alkanut lokakuussa 2018 ja jatkuu vuoteen 2021 asti. Heidän ratkaisussaan digitaalista tietoa tallennetaan 35 mm filmille, joka kestää heidän mukaansa aikaa paljon paremmin kuin muut tallennusratkaisut. Perusmigraatioissa on heidän mielestään



Digitalian VR ratkaisua demoamassa Teemu Hänninen. Ruutukaappaus Digitalian esittelyvideosta / Darcmmedia.

aina ongelmia – toisaalta Digitalian tekemissä migraatioissa ei ole muutamia jo ratkaistuja enkoodausongelmia lukuun ottamatta esiintynyt mitään ongelmia. iVM:n ratkaisuna on filmirullalle tallennettu konekielinen tieto sekä dekodeausohjelma virtuaalikoneessa, joka käsittelee datan. Jokaisella filmirullalla on mukana hakemisto, ohjeet 35 mm filmiskannerin rakentamiseen, kuinka kääntää / rakentaa virtuaalikone, alustusohjelma, pääohjelma sekä varsinainen digitaalinen sisältö. Näin äkkiseltään vaikuttaa huomattavasti monimutkaisemmalta kuin perinteisen migraation tekeminen aika ajoin, mutta aika näyttää ottaako tämä ratkaisu tuulta siipiensä alle vai ei. Tällä hetkellä ratkaisu on tarkoitettu multimediasisällön säilyttämiseen, mutta sisarprojektissa (VirtuMa) tarkoitus tukea myös PDF/A-formaattia.

Seuraavaksi Mats Porsklev, Sydarkiveralta Ruotsista esitteli kunnille tarkoitettua arkistoratkaisua, jossa on tällä hetkellä mukana 27 kuntaa. Kyseessä on perusarkistoratkaisu, kahdennetulla säilytyksellä, joista toinen on vuoren sisällä oleva pimeä arkisto. Maksupolitiikkana heillä toimii kunnan asukasluku. Tämä ei kuulemma ole kovinkaan helppo yhtälö, kun otetaan huomioon se, että Sydarkivera toimii viranomaisstatuksella, mikä tarkoittaa, että he eivät saa tuottaa voittoa. Siksi maksun tulee olla balanssissa kustannusten kanssa. Viimeisenä Tšekin kansallisen digitaalisen arkiston Pavlína Nimrichtrová kertoi heidän toiminnastaan, lainsäädännön vaikutuksista sekä Archivematican¹ käytöstään.

Lukas Rosenthaler Baselin yliopistolta puhui keynote-puheenvuorossaan tutkimusdatan säilyttämisestä

¹ <https://www.archivematica.org/en/>

FAIR (Findable, Accessible, Interoperable, Reusable) -periaatteiden mukaisesti. FAIR:n mukaan kaikki tutkimustieto on oltava arkistoituna ja uudelleen käytettävissä sekä kenen tahansa tutkijan hyödynnettävissä. Tämäkään ei ole helppo toteuttaa, koska heilläkin on käytössään monimuotoisia tutkimustietokantoja, myös kotikutoisia.

Lisäksi Rosenthalerin mielestä myös tietokantoihin tehtävät kyselyt pitäisi pitää tallessa, koska se kertoo, mitä tutkija on kannasta etsinyt ja miten. Kaiken tuuppaaminen OAIS-paketteihin ei myöskään toimi, koska näiden pakettien käsittelyyn ei ole vielä yksinkertaisia työkaluja. Tähän on kehitetty DaSCH (Data and Service Center for Humanities) -pitkäaikaissäilytyspaikka ihmistieteisiin liittyen. Mitä ilmeisemmin tämä ratkaisu on vastaava kuin Yhteiskuntatieteellinen tietoarkisto. DaSCH-palveluun kuuluu tallennustilan lisäksi myös työkaluja muun muassa tiedon uudelleen käytön helpottamiseksi. Lähtökohtaisesti kaikki on suunniteltu niin, etteivät tekniset käyttäjät osaa hyödyntää sisältöä tutkimuksessaan. Ratkaisu perustuu RDF:ään, Triplestoreen sekä avoimeen linkitettyyn tietoon.

Rosenthaler totesi, että yleisin tapaus yliopistoilla tehtävässä tutkimuksessa on edelleen se, että projekti ja rahat loppuvat ja data jää makaamaan johonkin servereille tai levyille. Sitten optimaaltilanteessa se otetaan tavalla tai toisella talteen. Esimerkkinä hän mainitsi kreikkalaiseen mytologiaan liittyneen projektin, joka oli kestänyt 25 vuotta. Projektin tietokannan aikanaan toteuttanut firma oli mennyt konkurssiin 15 vuotta sitten ja koko kanta piti takaisinmallintaa (reverse enginee-

ring). Pelkkä takaisinmallinnus oli vienynt noin kaksi vuotta, mutta nyt tiedot ovat netissä kaikkien tutkijoiden käytettävissä. Kysymyksissä kävi ilmi, että kehitetty DASCH-ratkaisu on avoimesti Githubissa saatavilla, mutta se on edelleen kehitteillä. Rosenthaler myös totesi, että ratkaisu on pääasiassa Sveitsin tutkimuksen käyttöön, mutta ulkomaisista tallenteista voi neuvotella, jos hinnasta päästään sopuun.

Seuraavan session anti oli ainakin allekirjoittaneelle turha. Slovenian Mikrografija-yrityksen Sandi Rozman puhui siitä, kuinka arkistointia koskeva lainsäädäntö vaikuttaa järjestelmätoimittajien toimintaan. Sveitsin Ecma Internationalin István Sebestyén ehdotti 500–1000 vuotta kestäviä medioita ja jpeg-kuville OAIS-standardin kaltaista omaa standardia. Huomioitavaa tässä esityksessä oli suomalaisen Tomi Ahonen consulting -firman materiaaleista napatut kuvat. Session viimeisissä puheenvuoroissa Sonia Renada UK:n kansallisarkistosta puhui heidän Bayesialaisen riskimallin päivityksestä ja Karin Bredenberg kertoi DLM-Forum DILCIS (Digital Information LifeCycle Interoperability Standards Board) kuulumisia.

Viimeinen keynote oli kiinnostava. Sarah Kenderdine (École polytechnique fédérale de Lausanne) puhui uudenkaltaisista käyttöliittymistä, joilla kulttuurillista dataa voidaan esittää ja muokata. Hän on ollut mukana toteuttamassa noin 80 erilaista museoinstallaatiota ympäri maailmaa. Konseptina on ollut EM+ experimental museum. Moni esitelty toteutus oli ns. live scale-periaatteella eli käytössä oli joko kiinteitä rakennelmia tai kymmeniä suuria ruutuja, joille esitettävät asiat

tuotiin. Yksi esimerkeistä on Pure Land: inside the Moga Grottoes at Dunhuang². Näin saa luonnollisesti näyttävän näköisiä esityksiä aikaiseksi, mutta kustannukset ovat huomattavasti suuremmat kuin esimerkiksi perus VR-tekniikkaa käytettäessä, vrt. 20 kpl 65" 4k ruutua vs. pelitietokone ja VR lasit (20*1000 € = 20 000 € vs. pelikone 3000 € + Pimax 8k VR lasit 1300 € = 4300 €), sähkönkulutuksesta puhumattakaan. Hyvä ja kiinnostava aihe, mutta lässähti osittain esityksen lukemiseen paperilta. Pääpointtina oli kuitenkin se, että nykyaikana digitaaliset museo-objektit eivät toimi enää vain oikeiden objektien tukena, vaan ne ovat saavuttaneet oman osansa itsenäisinä museoesineinä

Tunnin mittaisessa eArchiving sessiossa käsiteltiin E-ARK4ALL-hankkeen tuloksia. Jan totesi, että eArchivingin saaminen mukaan EU:n CEF building blockiin oli todella suuri etu koko yhteisölle. Hanke loppui 31.5.2019 ja kaikki toimenpiteet on saatu tehtyä onnistuneesti, muun muassa palvelutiski toimii osoitteessa <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eArchiving+Service+Desk>. Sinne voi laittaa kysymyksiä ja pyyntöjä liittyen CEF:n sähköiseen arkistointiin.

Viimeisen session aloitti Hanno Vares Viron kansallisarkistosta (NAE). Hän puhui aineistokuvausten uudelleenkäytöstä arkistoonvienti-vaiheessa. Tätä varten Virossa on kehitetty viimeiset kolme vuotta ASTRA nimistä palvelua, jossa arkistotietoa tuottavat toimijat voivat kehittää ja hyväksyttää NAE:lla omia luokitteluskeemojaan. Ennalta määritellyjä luokitteluja voidaan

² <http://alive.scm.cityu.edu.hk/projects/alive/pure-land-ii-2012/>

uudelleen käyttää arkistoon viennin yhteydessä. Tarkoituksena on tehdä arkistoon siirto tiedon tuottajille niin helpoksi kuin mahdollista.

Seuraavana Mette van Essen Hollannin kansallisarkistosta puhui sähköpostien automaattisesta luokittelusta. Suhteellisen tarkkoihin tuloksiin oli päästy opettamalla kahden kuukauden ajan valvottua koneoppimisjärjestelmää, jonka jälkeen järjestelmä oli otettu itsenäiseen käyttöön. Käyttäjän tulisi vain ajoittain tarkistaa, onko järjestelmä luokitellut sähköpostin oikein vai ei, ja korjata tarvittaessa väärä luokitus. Korjauksista menee luonnollisesti aina tieto järjestelmään, joten sen pitäisi oppia korjauksista. Lopuksi Mette totesi, että algoritmi tekee virheitä, mutta niin tekee ihminenkin. Algoritmin etuna on se, että se toimii aina samoin, kun taas ihminen on alttiimpi mielialan ja ympäristön muutoksille.

Päivän viimeinen esitys oli varattu johtajistoon valitulle Markus Merenmiehelle. Hän puhui rakenteellisen tiedon säilyttämisestä ja metatiedottamisesta. Hän esitti monia aiheellisia kysymyksiä liittyen tietokantoihin. Tiedetäänkö tietokannasta, mikä sen alkuperäinen tarkoitus oli, mitä se loogisia elementtejä se sisälsi, kuka tietokantaa käytti ja miten ja kuka on vastannut mistäkin asiasta. Jo GDPR:n mukaan sinun tulee tuntea tietosi eli pelkän tietokannan säilyttäminen ei vain riitä. Päivän päätti Merenmiehen kutsu kylmään ja märkään Helsinkiin 14.–15. marraskuuta 2019. Teemana DLM-Forumissa on silloin "trusted repositories". Nyt siis kaikki asiasta kiinnostuneet paikan päälle, koska etäisyys ja matkakulutkaan eivät ole enää esteenä.