

# Archiving 2019



Tuomo  
Räisänen  
It-asiantuntija  
Xamk  
Digitalia

**A**rchiving 2019 järjestettiin toukokuussa aurinkoisessa Lissabonissa. Osallistujia konferenssiin oli ilmoittautunut 166, tasaisesti ympäri maailmaa. Allekirjoittanut oli ainoa suomalainen tapahtumassa. Tapahtumapaikkana oli Portugalin kansallisarkisto. Esitykset muodostuivat esitelmistä ja posterisessioista, jollaisen allekirjoittanut piti.

Pääteemoja oli tänä vuonna kaksi. Ensimmäinen aihepiiri sisälsi valokuvien ja äänitteiden digitoimista ja esineiden kuvantamista digitaaliseen muotoon, ottaen huomioon pitkäaikais säilytyksen. Nykyään aineisto on jo lähtökohtaisesti digitaalista ja tältä osin käsiteltiin myös digikuvien ISO-standardeja. Toinen kokonaisuus muodostui massadigitoinnista ja tähän liittyen digitaalisten arkistojen muodostamisesta, käytettävyydestä ja tallennusratkaisuista.

Valokuvaaminen ja kuvien pitkäaikais säilytys, sisältäen laadun varmistuksen, kehittyvät teknisesti koko ajan. Konferenssi alkoikin katsauksella isäntämaan valokuvaarkistojen nykytilanteeseen. Portugalissa eräässä kuva-arkistossa on menossa jo kolmas digitointier-

ros. Arquivo Muncial de Lisboa on digitaalinen tallennus aloitettu jo vuonna 1992. Kuvien tarkkuus alkaa olla nykyisin sitä luokkaa, että tekniikan kehityksestä ei ole enää olettavissa hyötyä.

3D-kuvantaminen pienistä (1 cm) esineistä aina katedraaleihin asti oli myöskin vahvasti läsnä. Ei muodinkaan tallentamista ollut unohdettu. Kuvien koko on kasvanut kuitenkin varsin suureksi, esim. yksi puku 128 Gb. Tämä aiheuttaa sitten ongelmia käytettävyyden suhteen, tuollaista datamäärää ei nykyisillä web-tekniikoilla oikein selaimessa pyöritellä. Käytettävyys kuitenkin mahdollistuu ns. RTI (Reflection Transformation Imaging) -tekniikoilla, joissa kappaleen pinta pystytään esittämään tehokkaasti ja riittävällä tarkkuudella selaimessa. Tästä piti oivallisen esityksen Peter Fornaro, Digital Humanities Lab, University of Basel, Switzerland.

Massadigitointi oli kovasti tapetilla. Esitelyjen projektien koot vaihtelivat YK:n sadasta hyllykilometristä, muutaman sadan tuhannen sivun arkistoihin. Skannerit ovat nykyisin varsin kehittyneitä ja digitaalista aineistoa saadaan kyllä muodostettua rivakkaan tahtiin.

Useissa esimerkeissä kävi ilmi, että aineiston käytettävyyden takaminen tämän vauhdin keskellä on syytä miettiä tarkasti. Varmin tapa on kerätä metadata ihmisvoimin pala kerrallaan. Tämä tapa on myöskin hitain ja kallein, tosin käsikirjoitettujen aineistojen kohdalla

usein välttämättömyys, vaikka HTR (Handwritten Text Recognition) -menetelmät kehittyvät koko ajan. Suomessa HTR-menetelmiä kokeillaan READ-hankkeessa. Jos aineistoa voidaan tulkita OCR:n (Optical Character Reading) avulla, niin kuin painetulle tekstille voidaan jo hyvällä menestyksellä tehdä, päästään tunnistamaan tekstiä automaattisesti. Tätä voidaan sitten käyttää apuna metadatan luomisessa.

Käytännön esimerkin automaattisesta metadatan keräämisestä esitteli Martijn van der Kaaij, Heron Information Management LLP, Weert, The Netherlands. Metadatan tärkeys korostuu aineiston käytettävyyttä mietittäessä. Mitä parempi kuvaus aineistosta on, sen helpompi on eri alojen toimijoiden suorittaa aineistosta hakuja.

Massadigitointiin varatut resurssit, lähtötilanteet ja olosuhteet vaihtelevat kovasti. Optimitapaushan on hyvin inventoitu arkisto ja laboratorio-olosuhteet. Maailmalla on kuitenkin projekteja, joissa lähtökohtana on monikansallinen projekti, inventoimaton arkisto, työtilana entinen vankila, olosuhteet kuumat ja kosteat, henkilökunnalla ei yhteistä kieltä, sähkön saanti on puutteellista ja IT-osaaminen vajavaista. Projektien kustannustehokkuuksien vertailu on näin ollen kohtalaisen hankalaa.

Esimerkinomaisesti: YK:n arkistoissa noin 2,7 kilometrin aineiston tallentamiseen on varattu aikaa 5 vuotta ja 25 henkilöä, League of Nations

Archives (Geneve), toteuttaa työn. Kambodžassa taasen n. 400 000 sivun (muutama kymmenen hyllymetriä) digitointiin on varattu kymmeniä henkilötyövuosia.

Arkistoitava materiaali on nykyään jo alun perin digitaalisessa muodossa. Workflow on kuitenkin samankaltainen massadigitointiin verrattuna. Käytettävyyden takaamiseksi metadata on luotava tavalla tai toisella. Esimerkkinä tästä on EU:ssa meneillään oleva digitaalisen aineiston säilöntäprojekti, sen

toteuttaa Publication Office of the European Union Luxemburg, muiden tehtäviensä ohessa. Aineistoa kuvaavan datan kerääminen on heilläkin ongelma, mutta osa tästä työstä on voitu automatisoida, aivan kuten meillä Digitaliassa.

Myöskin tekniikan alalla mietitään tiedon arkistointia. CERNissä, hiukkaskiihdyttimen äärellä, tuotetaan valtava määrä mittausdataa, joka on tarkoitus arkistoida. Maailman suurimman tieteellisen infrastruktuurin digitaalisen muistin projektipäällik-

kö Jean-Yves Meur kertoi, että datan määrä on muutamia petatavuja sekunnin murto-osassa. Aineiston koosta saa kuvan, kun ottaa huomioon, että CERNissä on käytössä noin 1500 palvelinkonetta. Eikä tarkoitus ole säilöä pelkkää tieteellistä dataa, vaan aineisto on sangen monipuolista, sisältäen muun lisäksi kokouksien äänitteitä, valokuvia, videoita, tieteellisiä julkaisuja. Kaikkea tietoa ei pystytä mitenkään tallentamaan pitkäaikaisesti vaan seulontapäätöksiä on tehtävä.

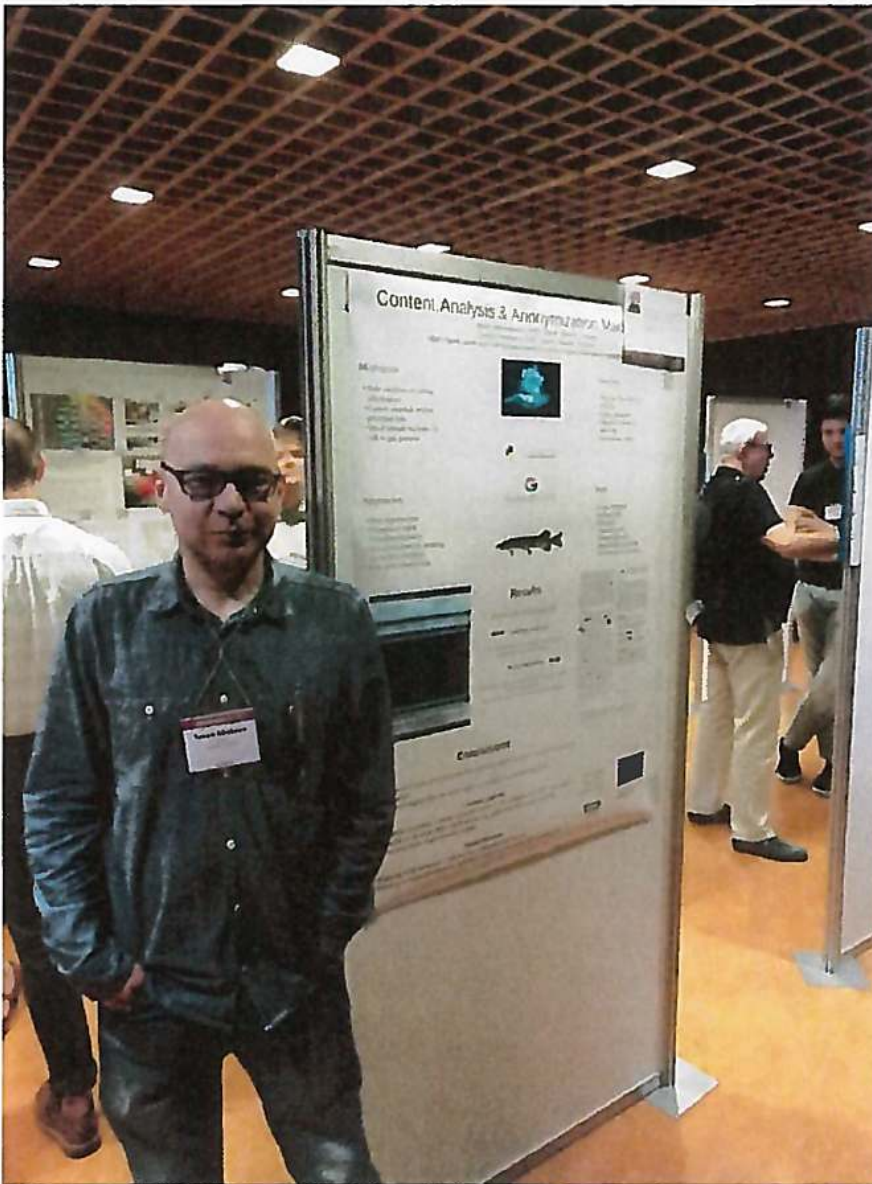
Avoin datan saatavuus ja käytettävyys ovat tärkeitä yhteisiä teemoja digitaalisissa arkistoissa. Web-pohjaisia ratkaisuja esittelivät useat toimijat. Käymällä eri arkistonmuodostajien sivuilla voi seurata digiarkistojen käytettävyyden kehitystä. Yhteenvetona digitaalisen säilytyksen workflowsta voi helposti todeta, että vaikka työn vaiheet ovatkin aina suhteellisen standardeja, toteutus on aina tapauskohtaisesti erilainen.

Digiarkistojen pitkäaikais säilytykseen mielenkiintoisen vaihtoehdon esitteli Morten Thorkildsen Norjan kansallismuseosta. Arktinen maailman arkisto sijaitsee Huippuvuorilla, jossa aineistojen vartioimisesta huolehtivat mm. jääkarhut. Aineisto on tallennettu ei-magneettiselle filmille ja näille keloille luvataan muutaman sadan vuoden elinikä. Kelat sijoitetaan lukulaitteineen entisiin kaivoksiin.

USA:ssa ei ehkä virallisesti tunnusteta ilmastonmuutosta, mutta NASA kuitenkin varautuu pahan päivän varalle. Digitaalista Nooan Arkkia (DNA-arkki) testataan parhaillaan kansainvälisellä avaruusasemalla. Write Once Read Forever (WORM) -tallennusmenetelmän ei pitäisi olla herkkä avaruuden



Mäkiä riittää Lissabonissa.



Postereille oli varattu reilusti tilaa.

taustasäteilylle. Richard J. Solomon (University of Pennsylvania, USA) työryhmineen on kehittänyt laitteen. Tarkoituksena on tallentaa maapallon DNA:ta kattavasti. Tilantarve on arvioitu jottatavuuissa (1 perässään 24 nollaa). Laite muodostuu millimetrin paksuisista "siivuista". Tallennuskapasiteetti 4x5 tuuman levyllä on tällä hetkellä yli neljä teratavua.

Digitalian aiheena oli tällä kertaa digitaalisen aineiston anonymisointi ja yleisemmin sisältöanalyysi. Viimeistään posterisession aikana tuli selväksi, että aihe on ajankohtainen

muuallakin. Kyselyitä käytetyistä tekniikoista tuli kohtalaisesti. Myöskin mahdollisia jatkokehitysideoita esitettiin. Kaiken kaikkiaan konferenssin ilmapiiri oli rento, ja siitä pitivät osaltaan huolen konferenssin jouhevät järjestelyt.

Vapaa-ajanohjelmaan ei juurikaan jäänyt aikaa, vaikka Lissabonissa nähtävää olisi riittänyt. Itse tein muutaman tunnin vapaa-ajalla minirailin Lissabonin metrolla, joka onkin ylivoimaisesti näppärin kulkuväline ruuhkaisessa kaupungissa.

## FAILI

Faili on Liikearkistoyhdistyksen julkaisema ainoa suomalainen asiakirja- ja tiedonhallinnan sekä arkistoinnin ammattilehti.

Faili kertoo neljästi vuodessa monipuolisesti alan koti- ja ulkomaisista tapahtumista, ihmisistä ja ilmiöistä sekä tutkimuksesta ja koulutuksesta.

Liikearkistoyhdistyksen jäsenille Faili sisältyy jäsenmaksuun, yhteisöjäsenmaksuun sisältyy kaksi Failin vuosikertaa.

Failin tilaushinta on 66 € ja päätoimisille opiskelijoille 16,50 €.

Arkistoalan yhdistysten jäsenten (AY, KAY, TAAJ) alennushinta on 33 €/vuosi.

Failissa julkaistava aineisto on toimitettava päätoimittajalle viimeistään:

- numero 4/2019, 8.11.2019
- numero 1/2020, 21.2.2020
- numero 2/2020, 8.5.2020
- numero 3/2020, 28.8.2020
- numero 4/2020, 6.11.2020

Päätoimittajana toimii Juha Henriksson, juha@musiikkiarkisto.fi, 045 249 5680.