

Nro **3**/2015 ← [Takaisin etusivulle](#)



Mikkeliin perustettu tutkimuskeskus Digitalia tuo laajat tietoaaineistot hyötykäyttöön

Uusi tutkimus- ja kehittämiskeskus Digitalia vahvistaa suomalaista digitaalisen tiedonhallinnan osaamista. Keskus edistää digitaalisten aineistojen käytettävyyttä, avaamista, louhimista, saatavuutta ja hyödyntämismahdollisuuksia.

Mikkelin ammattikorkeakoulu ja Kansalliskirjaston Digitointi- ja konservointikeskus ovat 2000-luvun alusta asti tehneet tiedon ja sen pitkäaikaissäilyttämisen tutkimus- ja kehittämistoimintaa Mikkelissä.

Kesällä 2015 toimintansa aloittanut Digitalia syventää aiempaa yhteistyötä. Se tähtää osaksi laajempia kotimaisia ja kansainvälisiä digitaalisuuden ja tiedon tutkimuksen verkostoja.

Lisää käytettävyyttä digitaaliseen tietoon

Digitaliassa tutkitaan ja kehitetään digitaalisia aineistoja koskevia ratkaisuja, jotka hyödyttävät niin julkista sektoria, yrityksiä, muistiorganisaatioita kuin kansalaisten arkea.

Painopisteitä ovat tiedonhallinnan ja -keruun menetelmät ja työkalut sekä digitaalisen säilyttämisen menetelmät ja välineet. Keskeinen osa Digitaliassa tehtävää käytännön työtä on digitaalisten aineistojen käytettävyyden parantaminen, esimerkiksi etsimällä automaattisesti erilaisia sisältöjä.

Digitalia tutkii ja soveltaa älykkäitä ja oppivia ratkaisuja, joiden avulla voidaan käsitellä suuria datamassoja.

Data on vasta raaka-ainetta

Maailmassa syntyy joka hetki valtavia määriä dataa. Kansalaisen arjessa tyypillisiä esimerkkejä ovat maksutapahtumat kauppojen kassoilla ja aktiivisuusrannekkeiden tekemät mittaukset. Kuitenkin vain pieni osa datasta on informaatioisällöltään ainutlaatuista, sillä eri järjestelmiin tallentuu myös päällekkäisiä tietoja.

Haasteena on, että datasta tulisi tunnistaa ihmisten toiminnan kannalta mielekkäät ja hyödylliset osat. Data onkin ensin muutettava ymmärrettäväksi ja hyvin kuvailluksi tiedoksi, jonka alkuperä, merkitys ja ominaisuudet tunnetaan.

Digitalia kehittää menetelmiä tiedon automaattiseen keräämiseen, luokitteluun ja rikastamiseen. Lisäksi kartoitetaan tapoja, joilla kansalaiset ja yritykset voivat paremmin hallita omaa digitaalista tietoaan ja yhdistää näitä muihin rekistereihin.

Arvokas tieto esiin massasta

Tärkeitä konkreettisia pilottiaineistoja Digitaliassa ovat digitaalisen viestinnän arkistot, digitoidut suomalaiset sanomalehdet ja julkisen sektorin päätösasiakirjat. Tavoitteena on, että kehitettävät menetelmät ovat hyödynnettävissä myös muilla vastaavilla aineistoilla.

Laajoista sähköpostiarkistoista seulotaan esiin olennainen tieto, digitaaliset pdf-asiakirjat jaetaan pienempiin osiin asiasisällön perusteella ja mahdollistetaan henkilötietojen anonymisointi. Asiakirjoihin lisätään tunniste- ja kuvailutietoja, jotka helpottavat aineiston jatkokäyttöä.

Sanomalehtiaineiston käytettävyyttä parannetaan Kansalliskirjaston johdolla koneoppimisen menetelmin. Artikkeleita voidaan poimia automaattisesti tekstimassasta ja mahdollisesti myös tehdä tiivistelmiä.

Digitaliassa tehdään digitaalisen humanismin tutkimusta yhteistyössä mm. historian tutkimuksen kanssa ja luodaan uusia tapoja seuloa uutta tietoa aikaisemmin jäsentelemättömästä aineistosta. Samalla yhdistely toisiin aineistoihin ja muihin konteksteihin helpottuu.

Tiedon koko elinkaari huomioon

**Maailmassa
syntyy joka
hetki
valtavia
määriä
dataa.**

**Digitalisointi
ei ole niin
yksinkertaista
kuin usein
oletetaan.**

Digitalisointi ei ole niin yksinkertaista kuin usein oletetaan. Uusien teknologisten ratkaisujen innoittamana saattaa unohtua tiedon pitkäaikaissäilytyksen tarve ja sen todelliset kustannukset. Digitaalisen viestinnän aikajänne on tyypillisesti lyhyt – esimerkiksi Twitter-viestin keskimääräinen vaikuttavuusaika lasketaan minuuteissa.

Toisessa ääripäässä ovat viralliset asiakirjat, joiden säilytysaika mitataan kymmenissä vuosissa, ja kulttuurihistoriallisesti merkittävät aineistot, jotka säilytetään ikuisesti.

Digitaalisen pitkäaikaissäilytyksen kehittäminen edellyttääkin erilaisten säilytysmedioiden, tiedostomuotojen ja säilytysalustojen testausta. Digitalia edistää digitaalisen tiedon siirtämistä vanhentuneilta alustoilta ja medioilta uusiin. Pilottiaineistoilla tehtävä automaattisten sisältökuvausten tuottaminen tukee myös erilaisten asiakirjojen hallintaa.

Digitalia

- Mikkeliin on perustettu digitaalisen tiedonhallinnan tutkimuskeskus Digitalia
- Digitalian muodostavat Mamk, Helsingin yliopisto ja Kansalliskirjaston Digitointi- ja konservointikeskus, lisäksi useita yhteistyökumppaneita
- Tutkimuskeskus on perustettu hankkeessa, joka toteutetaan ajalla 1.5.2015–30.4.2017
- Hanketta rahoittaa Etelä-Savon maakuntaliitto Euroopan aluekehitysrahastosta
- Verkkosivut: <http://www.digitalia.fi>, Twitter-tunniste: **#DigitaliaFI**

Kirjoittanut Miia Kosonen ja Mikko Tolonen

Kirjoittajat työskentelevät Digitaliassa. KTT Miia Kosonen toimii TKI-asiantuntijana Mikkelin ammattikorkeakoulun sähkö- ja informaatiotekniikan laitoksella ja FT Mikko Tolonen professorina Helsingin yliopiston Kansalliskirjastossa.

AVAINSANAT:

arkistointi Digitalia tiedonhallinta

Artikkelit samasta aiheesta