

Uusi lataustyökalu digi.kansalliskirjasto.fi-aineistojen lataukseen

Liisa Näpärä
Suunnittelija

Digitaalinen avoin muisti (DAM) –hanke
Missä mennään –webinaari 14.9.2021



Vipuvoimaa
EU:lta
2014–2020



Digitalia
Digitaalisen tiedonhallinnan
tutkimus- ja kehittämiskeskus

Digitaalinen avoin muisti –hankkeen perustiedot

- Hankkeen kesto 1.10.2019 - 30.9.2021
- Euroopan aluekehitysrahaston rahoittama Vipuvoimaa EU:lta 2014-2020
- Yhteistyössä Kaakkois-Suomen ammattikorkeakoulun (Xamk) kanssa
- Kolme työpakettia
 - WP 1: **Käyttäjälähtöinen tieto** (KK ja Xamk)
 - WP 2: Tiedon visualisointi (Xamk)
 - WP 3: Sosiaalisen median arkistointi (Xamk)
- Huomio Kansalliskirjaston digitaalisissa aineistoissa ja niiden käytön kehittämisessä
 - Kysely, haastattelut, osallistuminen tutkimushankkeisiin
 - Digitaalisten aineistojen ympärille rakentuvien palveluiden kehittäminen kansainvälisten esimerkkien avulla

digi.kansalliskirjasto.fi-aineistojen lataustyökalu

- Esimerkki digitaalisten aineistojen käytön kehittämisestä käyttäjälähtöiseen tietoon perustuen
 - Kyselyssä toiveita digitaalisten aineistojen käytön helpottamiseksi
 - Idea- ja suunnitteluvaiheessa mukana erityisesti teknisesti orientoituneiden tutkijoiden haastatteluissa
 - Testausvaiheessa 7 käyttäjäpalautetta

digi.kansalliskirjasto.fi-aineistojen lataustyökalu

- Isojen tai tarkasti yksilöityjen aineistokokonaisuuksien lataamiseen digi.kansalliskirjasto.fi:sta
- Tekijänoikeudesta vapaille tai pääosin vapaille aineistoille
 - Sanomalehdet vuoden 1918 loppuun asti
 - Aikakauslehdet ja pienpainatteet vuoden 1910 loppuun asti
 - Muut aineistot ovat ladattavissa tekijänoikeuden metatiedon mukaan, jolloin aikaraja vaihtelee aineistoittain ja teoksittain

Lataustyökalun käyttö

- Työkalu on ladattavissa osoitteesta <https://digi.kansalliskirjasto.fi/opendata/submit>
 - Ladattaessa annettava lyhyt vapaamuotoinen kuvaus käyttötarkoituksesta esim. lataustyökalun kokeilu, tutkimuskäyttö tai sukututkimus
 - Työkalun mukana tulee käyttöohje
 - Samasta osoitteesta ladattavissa myös valmiit datapaketit
- Ohjevideo työkalun käyttöön <https://youtu.be/o-ykh9zTOdl>

Lataustyökalun käyttö

- Mahdollistaa Digistä haettujen aineistojen latauksen hakutulokset sisältävän sivun urlin perusteella
- Käyttäjä voi valita ladataanko tekstiaineisto ALTO xml tiedostoina vai tavallisina tekstitiedostoina. Sivukuvat myös ladattavissa.
- Lataus on mahdollistaa pysäyttää ja jatkaa myöhemmin loppuun
- Ladattavista aineistoista on tallennettavissa csv-formaatissa tarkempaa tietoa

Tekninen toteutus

- Toteutettu Pythonilla
 - Hyödyntää useita Pythonille yleisesti saatavilla olevia kirjastoja, käyttäliittymä tkinter kirjasto
- Työkalun käyttöön ei tarvita edistyneitä teknologisia taitoja, käyttäjä ei tarvitse Python-asennusta tai –kirjastoja
 - Kokeilunhalusta ei kuitenkaan haittaa alkuun pääsemiseksi
- Ladattua aineistoa voi analysoida eri tavoin omien tarpeiden mukaan
 - Myös hankkeessa myöhemmin julkaistavan päivitetyn Jupyter notebookin avulla

Lopuksi

- Julkaistiin kesällä - tiedotettiin elokuussa
- Tulossa kieliversiot ruotsiksi ja englanniksi
- Palaute ja kehitysideat ovat tervetulleita
- Tule kuulemaan lisää 23.9.2021 klo 16–17
 - https://www.lyyti.fi/reg/Digin_lataustyokalu_tutuksi

www.kansalliskirjasto.fi



Vipuvoimaa
EU:lta
2014–2020



Digitalia
Digitaalisen tiedonhallinnan
tutkimus- ja kehittämiskeskus