

INFORMAATIOTUTKIMUKSEN PÄIVÄT 2024



Ajankohtaisia näkökulmia datan tarjontaan Kansalliskirjastossa

Liisa Näpärä

Kansalliskirjasto

liisa.napara@helsinki.fi

<https://orcid.org/0000-0002-6473-5896>

Asiasanat: data, palvelut, tutkimusaineisto

Pysyvä osoite: <https://doi.org/10.23978/inf.148606>

Esityksessä kerrotaan Kansalliskirjaston datan tarjonnasta ja kehittyvistä datapalveluista tutkijoille. Tarjottavien palveluiden taustalla on 1) datan käyttö itsenäisesti, 2) koulutus ja opastaminen Kansalliskirjaston datan käyttöön 3) yhteistyö erilaisissa dataa tarvitsevilla tutkimus- ja kehityshankkeissa, erityisesti niissä tilanteissa, joissa olemassa olevat valmiit palvelut tai ratkaisut eivät riitä käyttäjälähtöiseen datan käyttöön. Erilaisten sisäisten ja ulkoisten kehittämishankkeiden tavoitteena on parantaa palvelujen laatua ja sijoittaa Kansalliskirjaston tarjoaman datan ja palvelut tutkimuksen linkkaaren oikeisiin vaiheisiin. Palveluiden kehittymistä tarkastellaan myös kohdentamalla kyselyjä tutkijoille, ja viimeisin näistä on tehty vuonna 2024. Lisäksi palautetta saadaan erilaisissa tilaisuuksissa ja yhteistyökuvioissa ja -tilaisuuksissa.

Tutkijan näkökulmasta Kansalliskirjaston (tai muiden kulttuuriperintöorganisaatioiden) tarjoama data näyttää usein tutkimusprosessin alkuvaiheessa, kun tiedon tarve kohdistuu kiinnostaviin datavarantoihin. Kansalliskirjaston tarjonnasta löytyy erilaisia mahdollisuuksia hyödyntää analogisia aineistoja, digitoituja kokoelmia, koneluettavaa dataa ja metadatta sekä sähkösyntyisiä sisältöjä. Aineistojen etsimisen voi aloittaa Kansalliskirjaston hakupalvelusta (kansalliskirjasto.finna.fi) tai hyödyntää verkkosivuille koottuja kuvauksia kirjaston kokoelmista ja datasta. Yhtenä haasteena on tutkimukselle tarjottavien sisältöjen löydettävyys ja aineistojen historiallisista syistä johtuva hajaantuneisuus eri palveluihin, mikä toisinaan hämärtää sekä tietoja aineistojen identifioimista Kansalliskirjastoon että kokonaisuudesta viestimistä tutkijoille. Kansalliskirjasto on digitoinut aineistojaan 1990-luvulta alkaen, ja tähän historiaan mahtuu monia eri käyttöliittymiä, mikä asiakkaan kannalta saattaa tuntua hämmentävältä. Digitoitu lehtiaineisto löytyy kattavasti digi.kansalliskirjasto.fi -palvelusta, mutta kirjoja on edelleen Digi-palvelun lisäksi myös Doria-julkaisuarkistossa. Lisäksi erityisaineistoille on omat palvelunsa Fragmenta membranea ja Fenno-Ugrica.

Vuoden 2024 Kansalliskirjaston digitaalisiin aineistoihin ja palveluihin tutkijoille suunnatun kyselyn mukaan digitoidut aineistot olivat varsin tuttuja, mutta koneluettavana tarjottava data oli tunnettua vain kolmannekselle vastaajista. Digitaalisten menetelmien käyttö kuitenkin vaatii, että koneluettavan datan tarjontaan ja sen sisältöön kiinnitetään aiempaa enemmän huomiota ja tästä on esimerkkinä digitointiohjelma, joka on luotu ohjaamaan kattavien kokonaisuuksien digitointia tuleville vuosille.

Tekijänoikeudesta vapaata dataa pääsee käyttämään datakatalogin (data.nationalibrary.fi) tai eri palveluiden yhteydessä olevien sisältöjen ja linkkien kautta. Haasteena on paitsi tekijänoikeudet mutta myös datapakettien pysyvyys, vaikka tarve versiointiin olisi sisältöjen kasvaessa ja muuttuessa. Muutaman viime vuoden aikana nopealla vauhdilla kehitynyt Lataustyökalu tarjoaa dynamisempaa sisältöä digitoitujen aineistojen käyttöön

tutkimuksessa. Parhaillaan tarkastellaan, voiko työkalun käyttötapoja laajentaa entisestään. Lisäksi käynnissä oleva humanistis-yhteiskunnallisten alojen Fin-Clariah-tutkimusinfrastruktuurihanke on jo tuonut ratkaisuja tekijänoikeudesta vapaan tekstimuotoisen datan käyttöön ja jatkossa parantaa myös visuaalisten aineistojen saatavuutta datana.

Usein tutkijoilla on laajemmat mahdollisuudet päästä käyttämään dataa kuin muilla käyttäjäryhmillä, sillä Kansalliskirjasto on lisensoinut erityisesti digitoituja aineistoja. Käyttöoikeudet mahdollistavat tällöin myös pääsyn tekijänoikeudenalaisten aineistojen louhintaan. Kaikkien aineistojen osalta sopimuskumppania ei ole ja tilanne on ongelmallisempi.

Osa dataan liittyvistä teknisistä ja oikeudellisista kysymyksistä vaatii toisinaan hyvin tiivistä yhteistyötä, jota toteutetaan erilaisissa tutkimusyhteistyöhankkeissa. Viime vuosina hankkeita on ollut useita, joissa on haettu ratkaisuja tekijänoikeudesta vapaan datan siirtämiseen laskentatehon lähelle sekä pilotoitu tutkimusaineiston takaisinottamista osaksi järjestelmiä. Vielä on kuitenkin paljon tehtävää sekä infrastruktuurin että yhteistyön näkökulmista, jotta nykyaikaisiin menetelmiin nojaava tutkimus on mahdollista täysin saumattomasti kaikenlaisilla tutkimukseen sopivilla aineistoilla.

Yhden haasteen asettaa myös datapuhe, jossa dataa on vähän kaikkialla ja eri tahot lähestyvät dataa eri tavoin. Toisille tutkimusdata on kaikkea mahdollista ja toisille se on nimenomaan vain tutkimuksessa käytettyä aineistoa. Tämä asettaa vaatimuksia sille, millaista tuki-infrastruktuuria datan käytölle luodaan toisaalta dataa tarjoavissa organisaatioissa ja yliopistojen datatuessa. Tämä vaikuttaa palveluun siitä, millä tasolla valmiiksi kuratoitua dataa tarjotaan ja miten tutkijat ja muut käyttäjät siihen suhtautuvat. Tutkimuksen elinkaaren eri vaiheissa oleva data on kuitenkin huomioitava laajasti ja dataan kiinnitetäänkin yhä enemmän huomiota eri suunnista sisällön, laadun, teknisten mahdollisuuksien sekä muun infrastruktuurikehityksen näkökulmista.