

28.10. aineistokatalogi-workshop ryhmätöiden yhteenveto

Mitä tietotarpeita käyttäjillä on aineistokatalogiin liittyen?

Tutkijat tarvitsevat mahdollisimman kattavia kuvauksia mahdollisimman monesta aineistosta. Tärkeää on tietysti että aineistokatalogi sisältää perustiedot aineistosta ja muuttujista.

Perustietoina pidettiin

- aineiston ja muuttujan nimi, ajankohta, perusjoukko, alue, lähde, kuvaus aineiston/muuttujan muodostamisesta, yhteystiedot

Perustietojen lisäksi ehdotettiin

- aineistotunnuksen luomista jokaiselle aineistolle,
- rinnakkaistietojen (melkein samaa tietoa sisältävät muuttajat) tuominen esille muuttujakuvauksessa, jotta tutkija voi valita mikä soveltuu parhaiten omaan tutkimusasetelmaan,
- tilastollisten tunnuslukujen ja jakaumatietojen ajamista jokaiselle muuttujalle,
- muuttujien käytettävyyttä ja laatua indikoiva pikaindeksi,
- linkkausta rekisteriselosteeseen,
- tietoja aineistoon ja muuttujiin kohdistuvista muutoksista vuosien varrelta (muutokset luokituksissa, keräystavoissa, mahdollisesti jopa lainsäädännössä)
- tietoa tietyn aineiston osajista eli aiemmista käyttäjistä ja aineistojen pohjalta tehdyistä julkaisusta -> Linkit julkaisuihin
- linkit aineistosta tehtyihin luotettavuustutkimuksiin ja
- johdettujen muuttujien käsittelyperiaatteiden/koodien tuomista esille.

Tutkijoiden mahdollista panosta aineistokatalogiin pidettiin tärkeänä. Tutkijat ovat usein keränneet aineistoista ja muuttujista tietoja useammasta lähteestä monien vuosien ajalta ja näitä olisi hyvä saada myös muiden käyttöön. Ideoitiin aineistokatalogiin erillistä 'tutkijoiden pöytälaatikko metatiedot'-osiota, jonne tutkijat voisivat tallentaa keräämänsä kuvaukset. Toivottiin myös tutkijoiden aktiivista osallistumista pikaindeksiosiossa (jossa tutkijat voivat käydä tykkäämässä tai ei tykkäämässä muuttujien käytettävyydestä/laadusta) sekä julkaisuluettelon ylläpidossa. Suunnitteella olevaa käyttölupaluetteloa sähköisessä lupahakemusjärjestelmässä tulisi käyttää hyväksi julkaisuluetteloa laadittaessa. Käyttöluvassa tulee olla ehto että tutkija ilmoittaa/vie julkaisutietonsa julkaisuluetteloon.

Muutosten hallintaa pidettiin hyvin tärkeänä. Aineistokatalogista tulisi selvittää vuosien varrella tapahtuneet muutokset luokituksissa sekä muuttujien ja aineistojen muodostamisessa. Lisäksi on hyvä jos laissa tapahtuneet muutokset, jotka vaikuttavat muuttujiin ja aineistoihin, ilmenevät kuvauksista.

Mitä toiminnallisuuksia käyttäjät tarvitsevat aineistokatalogissa?

Hakutoiminnot nähtiin aineistokatalogissa erittäin tärkeiksi. Ainakin seuraavat haut mainittiin: Boolean, algebra, poissulkeva (esimerkiksi älä näytä hakusanan X sisältäviä tuloksia), fasetti, vapaa teksti. Lisäksi mainittiin Muuttujakohtainen haku, joka toimisi sekä yhden aineiston, että koko aineistokatalogin sisällä.

Myös yhteys Kansallisarkistossa kehitteillä olevaan käyttöilupajärjestelmään nähtiin tärkeäksi. Sujuva siirtyminen säästäisi tutkijoilta aikaa ja vaivaa. Tämä toteutuisi varsinkin sitä kautta, että yhteen käyttöluvhakemukseen olisi yhdistettävissä monia eri aineistoja.

Palvelun monikielisyyden varmistaminen nähtiin myös tärkeänä. Varsinkin englanti nähtiin elintärkeänä kansainvälistä tutkimuskenttää ajatellen. Myös ruotsin nähtiin olevan tärkeä.

Myös aineistokatalogin yhteys dataan, eli etäkäyttöjärjestelmään, tulisi varmistaa. Tulisi välttää tilannetta, jossa käyttäjän tarvitsee mennä monen eri järjestelmän kautta päästäkseen käyttämään aineistoa.

Tutkijat ovat usein kiinnostuneita aineistojen havaintomääristä. Aineistojen etsimisen kannalta olisi erittäin hyödyllistä, että aineistojen selailun yhteydessä olisi aineiston kokonaishavaintomäärä näkyvissä. Näin tutkija voisi valita tutkimusasetelmaansa sopivan aineiston nopeammin ja helpommin.

Aineistokatalogissa tulisi olla avoin ohjelmallinen rajapinta, jonka kautta olisi mahdollista ladata aineistokuvauksia omaan käyttöön.

Olisi hyödyllistä jos aineistokatalogiin liitettäisiin aineistokohtainen ”lisätietolaari”. Sinne tutkijat voisivat viedä omia aineistoon liittyviä pöytälaatikkomateriaaleja, jotka ovat usein olemassa vain fyysisessä muodossa.

Koodikäsikirjojen saatavuus aineistokatalogissa on ehdottoman tärkeää. Niistä tulisi julkaista myös eri aikaversiot.

Olisi hyvä jos aineistokatalogissa olisi lisäksi julkaisuluettelo, joka sisältäisi myös DOI-linkit.

Lisäksi erilaisista rekistereistä linkitys muihin vastaaviin pohjoismaisiin aineistoihin nähtiin hyväksi ominaisuudeksi. Tämä edistäisi yhteispohjoismaisten tutkimusaineistojen käyttöä.

Missä muodossa organisaatioiden aineistot on kuvattu tällä hetkellä?

Aineistoja ei juurikaan ole kuvattu tai jos on, niin epäsystemaattisesti. Kun kuvauksia tarvitaan, niitä haalitaan eri lähteistä (esim. HILMO, tietojärjestelmien sisäänrakennetut metatietosetit, lomakkeiden täyttöohjeet, indikaattorien kuvaukset). Tietojärjestelmissä olevien kuvausten formaatit ovat usein peruskäyttäjälle vaikealukuisia. Standardoituja, yhtenäisiä kuvailumalleja ei käytetä mm. siksi, että halutaan pitää kiinni omista käytännöistä. Samassakin organisaatiossa kuvaustapoja saattaa olla monia. Kuvauksien säilytyspaikat ovat harvojen tiedossa eli niiden saatavuus on heikko. Tieto on yksittäisten henkilöiden hallussa, jopa hiljaisena tietona. Olemassa oleva kuvaustieto saattaa olla päällekkäistä, eli samoja tietoja on useammassa järjestelmässä ja tiedot eivät välttämättä ole ajan tasalla kaikissa järjestelmissä. Myös koodistojen kuvauksissa on puutteita, erityisesti vanhojen koodistojen osalta. Lisäksi koodistoja konvertoidaan väärin.

Systematisointia on kuitenkin tapahtumassa, ja asenneilmasto on muuttumassa myönteisemmäksi kuvauksille. Ymmärretään, että avoimen datan tavoitteet ja yhä lisääntyvä tiedon liikkuvuus edellyttää tiedon asianmukaista kuvaamista.

Erikseen mainittiin pohjoismaisessa syöpärekisterissä tekeillä oleva kuvausmalli, joka on lähiaikoina menossa käyttäjäarviointiin. Eläketurvakeskuksessa ja KELAssa on kuvauksia tehty jonkin verran. Kuvauksia ylläpidetään esim. Word-tiedostoissa.

Yhteiskuntatieteellisessä tietoarkistossa (FSD) kuvaukset tehdään DDI Codebook -tietomallin mukaisesti. DDI Lifecycle mahdollistaisi tiedossa tapahtuvien muutosten kuvaamisen Codebookia paremmin, mutta DDI-L on hyvin laaja, eikä tarvittavia työvälineitä sille ole saatavissa.

FSD:ssä on suunnitteilla työväline AILA-aineistokatalogin sisällöntuotantoon.