



TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

Metadatalu Tietoarkistossa

Tietoarkisto ja CESSDAn palvelut -webinaari
17.3.2021

Hannele Keckman-Koivuniemi



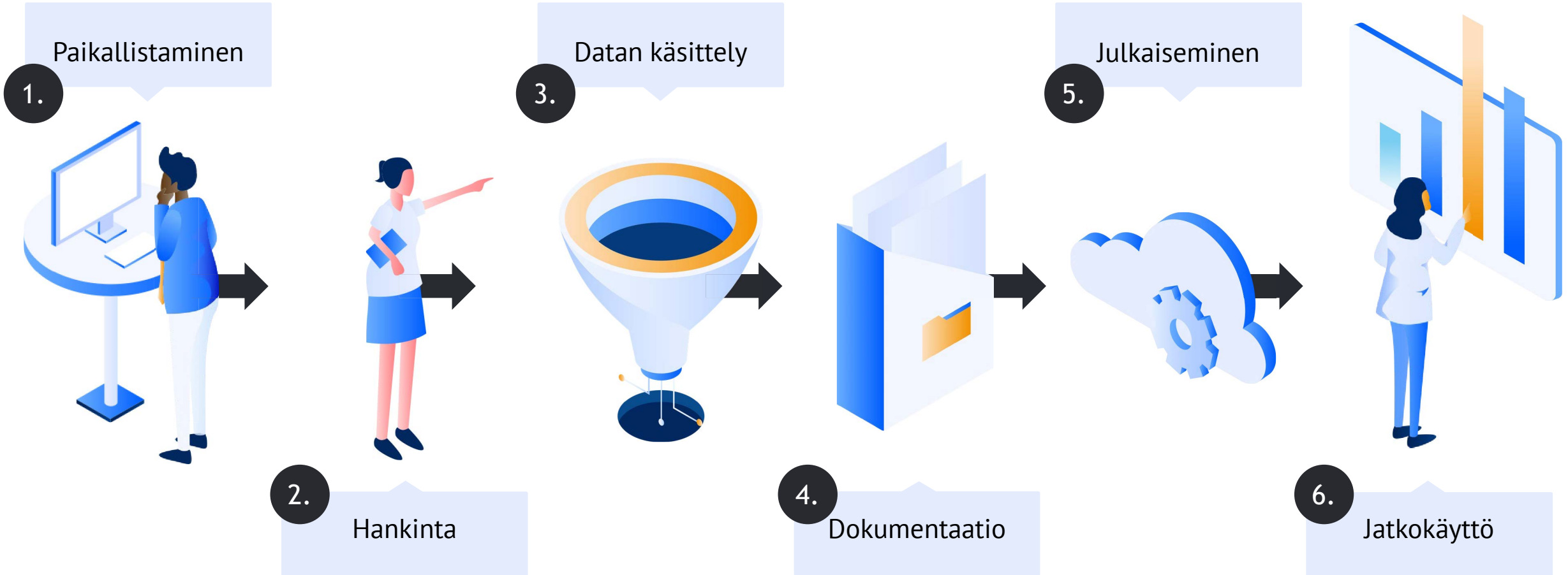
Tietoarkisto

Kansallinen tutkimusinfrastruktuuri, arkistoi ja välittää sähköisiä tutkimusaineistoja tutkimukseen, opetukseen ja opiskeluun

- ▶ Palvelut maksuttomia
- ▶ CESSDAn palveluntuottaja Suomessa
- ▶ CoreTrustSeal-sertifioitu
- ▶ Perustehtävinä
 - ▶ (ihmistieteellisten) aineistojen hankinta
 - ▶ hoivaaminen, säilytys ja toimittaminen jatkokäyttöön
 - ▶ tietopalvelu
 - ▶ vertailevan tutkimuksen edistäminen
 - ▶ vaikuttaminen (aineistot, avoimuus, parhaat käytännöt, standardit...)



Arkistointiprosessi



Järjestelmät

Sanastot esim.

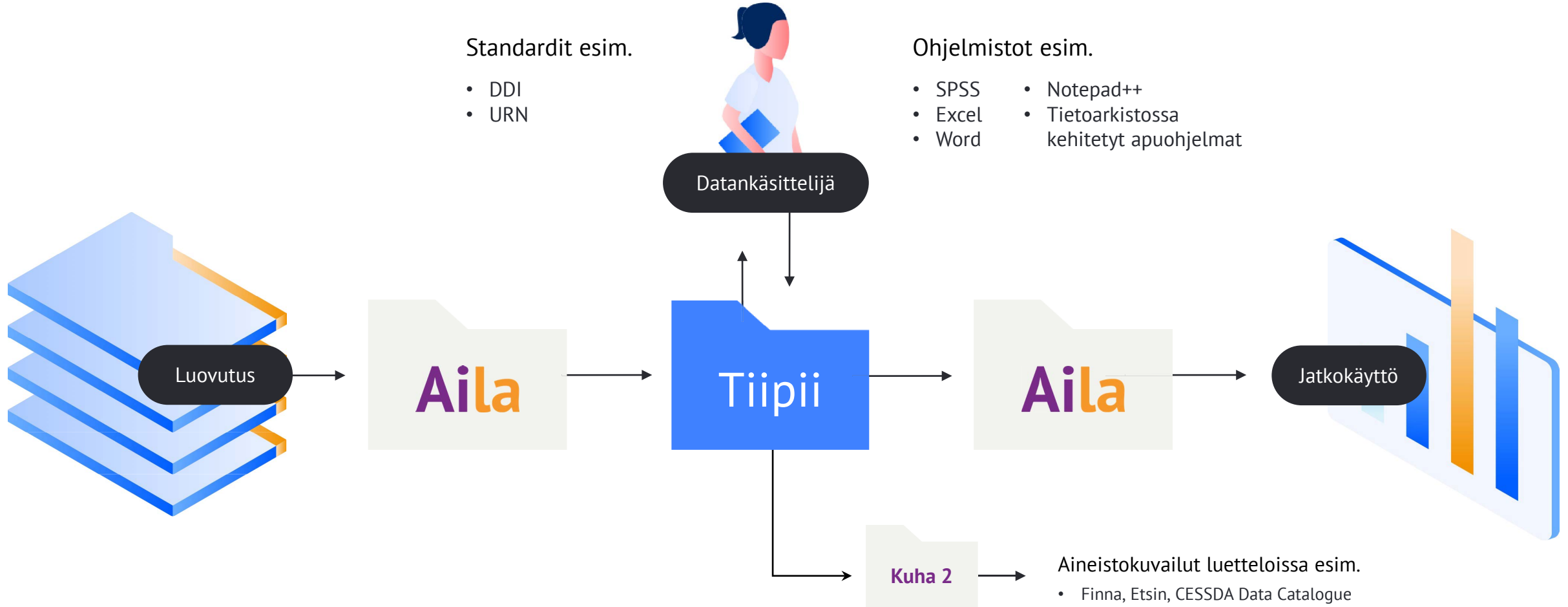
- YSO
- DDI CV
- CESSDA
- ELSST

Standardit esim.

- DDI
- URN

Ohjelmistot esim.

- SPSS
- Excel
- Word
- Notepad++
- Tietoarkistossa
kehitetyt apuohjelmat



Kuvailutyö Tietoarkistossa I

- ▶ **Tietoarkisto kuvailee** sinne tallennetut aineistot kansainvälisen, erityisesti yhteiskuntatieteellisten aineistojen tarpeisiin suunnitellun **DDI**-formaatin mukaisesti **XML**-kielellä sekä **sisältöettä muuttujatasolla**
- ▶ Sisällönkuvailut **suomeksi ja englanniksi**
- ▶ **Yksityiskohtainen kuvaileminen edellytys** sille, että **aineistojen pitkäaikaissäilytys ja jatkokäyttö** mahdollista ja aineistot **löytyvät helposti**
- ▶ mallin mukaiset XML-aineistokuvaukset mahdollistavat **tehokkaat haut** sekä tietojen **koneellisen julkaisun** eri muodoissa.
- ▶ Kv-kuvailumallin ansiosta aineistojen kuvailutiedot voidaan jakaa helposti myös moniin **kansallisiin ja kansainvälisiin yhteisluetteluihin.**

Kuvailut Tietoarkistossa: DDI-Codebook

Kolmisenataa tietoelementtiä jakautuvat viiteen osaan:

▶ Document Description

- ▶ mm. bibliografiset tiedot itse kuvailusta + lisenssitiedot



▶ Study Description

- ▶ Aineiston (ei siitä tehdyn tutkimuksen!) sisällönkuvailu, muun muassa tekijät, asiasanat, abstrakti, otannan kuvaus, tutkimusaineiston keruun kuvaus, havaintoyksiköt, perusjoukko, saatavuus, käyttöehdot

▶ Data Files Description

- ▶ tiedostojen kuvailu, muun muassa tiedoston rakenne, muuttujien määrä, tiedostojen koko, ohjelmistot

▶ Variable Description

- ▶ muuttujien kuvailu, esimerkiksi selitteet ja kysymystekstit

▶ Other Study-Related Material

Standardoituja sanastoja ja luokituksia käytössä

- ▶ Havainto/aineistoyksikkötyyppi
- ▶ Tutkimuksen aikaulottuvuus
- ▶ Otantamenetelmä
- ▶ Keruumenetelmä
- ▶ Keruuväline
- ▶ Tieteenala, OKM
- ▶ Asiasanat, suomenkieliset kuvailut, YSO
- ▶ Asiasanat, englanninkieliset kuvailut, ELSST
- ▶ Aihepiiri



Tutkimusaineiston kuvailu Tietoarkistossa kohta kohdalta

Tutkimusaineistojen kuvailussa käytetyt
kuvailuelementit.

Esimerkit pääosin FSD3513 ISSP 2020:
ympäristö IV: Suomen aineistosta,
<http://urn.fi/urn:nbn:fi:fsd:T-FSD3513>

AINEISTON PERUSTIEDOT

Aineiston nimi suomeksi ja englanniksi

Aineiston mahdollinen rinnakkainen nimi suomeksi ja englanniksi

Aineistonumero ja pysyvä urn-tunniste

```
<titlStmt>  
  <titl>ISSP 2020: ympäristö IV: Suomen aineisto</titl>  
  <parTitl xml:lang="en">ISSP 2020: Environment IV: Finnish Data</parTitl>  
  <IDNo agency="FSD">3513</IDNo>  
  <IDNo agency="Kansalliskirjasto">urn:nbn:fi:fsd:T-FSD3513</IDNo>  
</titlStmt>
```

TEKIJÄTIEDOT

Tekijät Tekijä = henkilö, yhteisö tai taho joka on vastuussa aineiston varsinaisesta ajatuksellisesta sisällöstä. Muodossa Sukunimi, Etunimi (aineistonkeruun aikainen taustaorganisaatio) tai Organisaatio

Muut tekijät Tahot, jotka osallistuneet esim. aineiston kokoamiseen tai käsittelyyn, mutta jotka eivät varsinaisia tekijöitä. Tällaisia mm. ne, jotka koodanneet, tallentaneet tai litteroineet aineistoa.

Aineiston tuottajat Organisaatio, jolla taloudellinen/hallinnollinen vastuu niistä fyysisistä prosesseista, joiden avulla aineisto on saatu aikaan. Esim. tilanne, jossa jokin organisaatio on tilannut aineiston tai ollut aloitteentekijä aineiston keräämiselle, mutta ei ole varsinaisesti tutkimuksen tekijä.

ARKISTOINNIN TIETOJA

Jakelija, Luovuttaja, Luovutuspv, Aineiston versionumero & Tietoarkiston aineistosarjaan kuuluminen

Malliviittaus

<biblit>International Social Survey Programme (ISSP) & Melin, Harri (Tampereen yliopisto) & Borg, Sami (Tampereen yliopisto): ISSP 2020: ympäristö IV: Suomen aineisto [sähköinen tietoaineisto]. Versio 1.0 (2021-02-19). Yhteiskuntatieteellinen tietoarkisto [jakaja]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD3513>
</biblit>

International Social Survey Programme (ISSP) & Melin, Harri (Tampereen yliopisto) & Borg, Sami (Tampereen yliopisto): ISSP 2020: ympäristö IV: Suomen aineisto [sähköinen tietoaineisto]. Versio 1.0 (2021-02-19). Yhteiskuntatieteellinen tietoarkisto [jakaja]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD3513>



SISÄLTÖ, KONTEKSTI JA VASTAAJAT

Asiasanat [YSO - Yleinen suomalainen ontologia](#) ja [ELSST - European Language Social Science Thesaurus](#)

Tieteenala/aihealue [Korkeakoulujen tutkimustiedonkeruussa käytettävä tieteenalaluokitus](#) pääotsikkotasolla & [CESSDAn aihepiiriluokitus](#)

Sisällönkuvaus Mahdollisimman tiivis yhteenveto, jossa kuvataan aineiston tarkoitus, luonne ja ala, sen sisällön erityispiirteet, tärkeimmät aihealueet sekä se, mihin kysymyksiin tutkimuksessa pyrittiin vastaamaan. Tärkeimpien kysymys/muuttujaryhmien luetteleminen on tässä olennaista. Abstraktissa pitäisi pyrkiä käyttämään termistöä monipuolisesti, jotta aineiston löytyvyys olisi mahdollisimman hyvä. Tulisi siis käyttää synonyymejä ja mahdollisuuksien mukaan ala- ja yläkäsitteitä. Mainitaan käytettyjä mittareita ja luetellaan taustamuuttujia. Tutkimushankkeen nimi ja rahoittajatieto.

Havainto/aineistoyksikkötyyppi Kohde, josta alkuperäisessä tutkimuksessa on tehty empiirisiä havaintoja tai jonka edustajana tutkittava on. Esim. henkilö, yksilö, ryhmä, organisaatio, valtio, kunta, tapahtuma, jäsenyys, lehtiartikkeli, TV-ohjelma jne. [CVS: DDI Controlled Vocabulary for Analysis Unit](#)

Perusjoukko/otos Kuvaus aineistonkeruun kohteena olleesta populaatiosta; siitä ihmis- tai asiaryhmästä, joka on tutkimuksen kohteena tai johon tutkimuksen tulokset viittaavat.

Aineiston laatu Karkea jako: kvantitatiivinen vai kvalitatiivinen



AIKA JA ALUE

Aineiston ajallinen kattavuus Aineiston tietojen käsittelemä ajanjakso

Aineistonkeruun ajankohta. Aineistonkeruun alkamis- ja loppumisajankohdat

Maa ja kohdealue

Tutkimuksen aikaulottuvuus Esim. poikkileikkausaineisto, pitkittäisaineisto. [CVS: DDI Controlled Vocabulary for Time Method](#)

```
<timePrd date="2020-00-00" event="single"/>
<collDate date="2020-09-21" event="start"/>
<collDate date="2020-12-22" event="end"/>
<nation abbr="FI">Suomi</nation>
<geogCover>Suomi</geogCover>
<timeMeth>Pitkittäisaineisto: trendi/toistuva poikkileikkausaineisto
  <concept>Longitudinal.TrendRepeatedCrossSection</concept>
</timeMeth>
```

KERUU JA OTANTA

Aineiston kerääjät Organisaatio tai Kerääjän Nimi (Organisaatio)

Otantamenetelmä/Havainto/aineistoyksiköiden valinta Miten tutkittavat on poimittu/valittu. Esim. kokonaisaineisto, satunnaisotanta, harkinnanvarainen poiminta. [CVS: DDI Alliance Controlled Vocabulary for Sampling Procedure](#) sekä sanallinen kuvaus otannasta/valinnasta

Keruumenetelmä [CVS: DDI Alliance Controlled Vocabulary for Mode Of Collection](#)

Keruväline [CVS: DDI Alliance Controlled Vocabulary for Type of Instrument](#)



AINEISTON SAATAVUUS JA KÄYTTÖ

* Aineistojen saatavuus:
(A) vapaasti käytettävissä ilman rekisteröitymistä (CC BY 4.0),
(B) tutkimukseen, opetukseen ja opiskeluun,
(C) vain tutkimukseen ja ylempiin opinnäytteisiin (esim. väitöstutkimukseen, pro graduun ja ylempään AMK-tutkinnon opinnäytetyöhön) ja
(D) vain luovuttajan luvalla.

Käyttöoikeudet Aineiston luovuttajan ja Tietoarkiston välisen arkistointisopimuksen ehdot aineiston jatkokäyttöön toimittamisesta*

Eriyisehdot Aineiston luovuttajan arkistointisopimukseen kirjaamat mahdolliset käytön lisäehdot.

Viittausvaatimus

Julkaisusta tiedottaminen

Varaumat

`<useStmt>`

`<restrctn>`Aineisto on käytettävissä (B) tutkimukseen, opetukseen ja opiskeluun.`</restrctn>`

`<citReq>`Aineistoon ja sen tekijöihin tulee viitata asianmukaisesti kaikissa julkaisuissa ja esityksissä, joissa aineistoa käytetään. Tietoarkiston antaman malliviittaustiedon voi merkitä lähdeluetteloon sellaisenaan tai sitä voi muokata julkaisun käytäntöjen mukaisesti.`</citReq>`

`<deposReq>`Tietoarkistoon on lähetettävä viitetiedot kaikista julkaisuista, joissa käyttäjä hyödyntää aineistoa.`</deposReq>`

`<disclaimer>`Aineiston alkuperäiset tekijät ja tietoarkisto eivät ole vastuussa aineiston jatkokäytössä tuotetuista tuloksista ja tulkinnoista.`</disclaimer>`

`</useStmt>`



MUUT TIEDOT

Lähdeaineistot

Painokertoimet Tiedot aineiston sisältämisestä painomuuttujista: nimet, miten ne on tehty ja miten niitä käytetään

Vastausprosentti

Aineiston käytössä huomioitavaa Aineiston mahdolliset puutteet, virheet ja poistot, aineistoon tunnisteteellisuussyistä tehdyt anonymisointitoimenpiteet ja muu tärkeä informaatio

Käytön ja kuvailun oheismateriaalit

Julkaisut Bibliografiset tiedot julkaisuista, joissa aineistoa on hyödynnetty

TIEDOSTOJEN KUVAILU

```
<fileDscr>  
  <fileTxt>  
    <fileName ID="FSD3513_file_1">daF3513.por</fileName>  
    <dimensns>  
      <caseQty>1137</caseQty>  
      <varQty>121</varQty>  
    </dimensns>  
    <fileType>SPSS Portable</fileType>  
  </fileTxt>  
</fileDscr
```

- Kvalitatiivisten aineistojen tiedostoja kuvataan sisällönkuvailun Aineiston määrä -kentässä



MUUTTUJIEN KUVAILU

Muuttujaryhmät:

Muuttujaryhmän yhteinen teksti (useimmiten kysymyspatterin teksti)

Muuttujat:

Nimi ja selite

Kysymysteksti (esiteksti, jälkiteksti ja varsinainen kysymysteksti) tai muuttujaa selittävä teksti

Haastattelijan ohjeet

Kuvailevat tunnusluvut ja suorat jakaumat

Arvot ja arvojen selitteet

Muuttujia ei asiasanoiteta. Tietoarkistossa on kehitteillä työkalu asiasanoituksen puoliautomatisointiin. Työkalu ehdottaa sopivia YSO- ja ELSST-asiasanoja (algoritmit).

```
<varGrp var="K35A K35B K35C K35D K35E  
K35F K35G">
```

```
<txt>Yleisesti ottaen, kuinka vaarallisia  
seuraavat asiat ovat mielestäsi  
ympäristölle? Valitse jokaiselta riviltä yksi  
vaihtoehto.</txt>
```

```
</varGrp>
```

```
<labl level="variable">[q1] Sukupuoli</labl>
```

```
<qstnLit>Mikä on sukupuolesi?</qstnLit>
```

```
<catgry>
```

```
<catValu>1</catValu>
```

```
<labl level="category">Mies</labl>
```


```
<catStat>500</catStat>
```

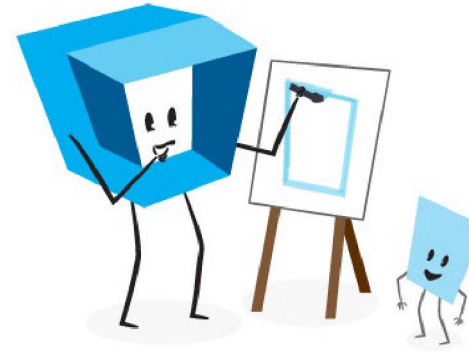
```
<catValu>2</catValu>
```

```
<labl level="category">Nainen</labl>
```

```
<catStat>500</catStat></catgry>
```

Kuvailutyö Tietoarkistossa II

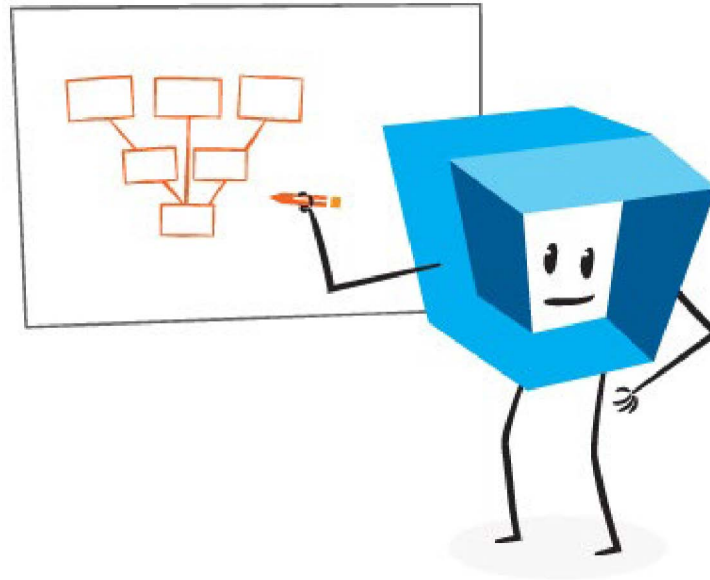
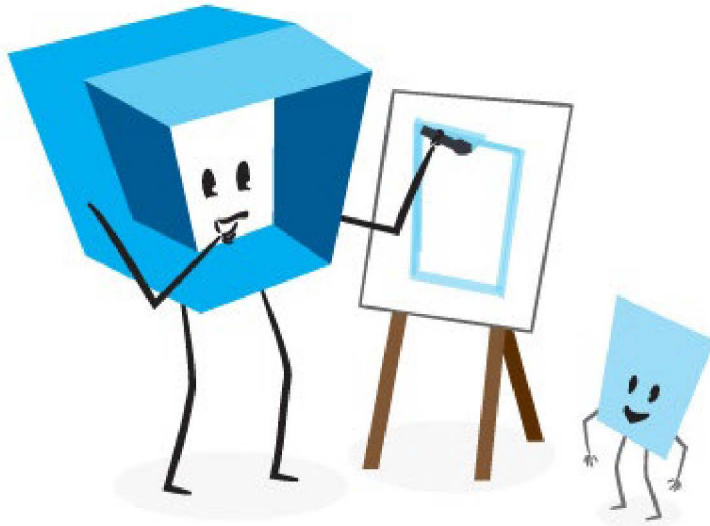
- ▶ Aineistomateriaalien luovutuksen yhteydessä **tutkija kirjaa Ailassa perustietoja** aineiston sisällöstä, keruusta ja aineistosta jo tuotetuista julkaisuista.
- ▶ Testiaineisto-DEMO: 
- ▶ Nämä tiedot tarvitsemme sisällöllisen metadatan tuottamiseen. **Muokkaamme ja lisäämme tietoja.**
- ▶ Muuttujakuvaailun pohjana Tietoarkistossa prosessoitu SPSS-datatiedosto.



Haluatko kuulla lisää?

- ▶ Näin Tietoarkisto palvelee arkistoinnissa
 - ▶ <https://www.fsd.tuni.fi/fi/palvelut/arkistointi/>
- ▶ Aineistohallinnan käsikirja: Aineiston kuvailu ja metadada
 - ▶ <https://www.fsd.tuni.fi/fi/palvelut/aineistohallinta/aineiston-kuvailu-ja-metadada/>
- ▶ CESSDA Data Management Expert Guide: Documentation and metadata
 - ▶ <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/Documentation-and-metadata>
- ▶ Kuinka olla enemmän FAIR? Tietoarkiston askeleita
 - ▶ <https://www.fsd.tuni.fi/fi/ajankohtaista/artikkelit/kuinka-olla-enemman-fair-tietoarkiston-askeleita/>

Kysymyksiä?



I love
Aila

Asiakaspalvelu
sähköposti: asiakaspalvelu.fsd@tuni.fi
puh. +358 29 452 0411



Jollei erikseen toisin mainita, tämä esitys on
lisensoitu Creative Commons Nimeä 4.0
Kansainvälinen -lisenssillä.

Hannele Keckman-Koivuniemi
hannele.keckman-koivuniemi@tuni.fi
ORCID: 0000-0002-9223-7261

www.fsd.tuni.fi