

# Sosiaalisen median aineistojen arkistoinnin mahdollisuudet ja mahdottomuudet

Sosiaalisen median aineistot tutkimuksessa –seminaari

17.11.2021 Tampereen yliopisto

Arja Kuula-Luumi

Johtava asiantuntija

Tietoarkisto

# Sisältö

- ▶ Tietoarkisto
- ▶ Somedatan arkistoinnin ongelmat
  - ▶ Etiikka ja Tietosuoja
  - ▶ Tekijänoikeudet
  - ▶ Käyttöehdot
- ▶ Aineistojen sisäänoton kriteereitä Tietoarkistossa

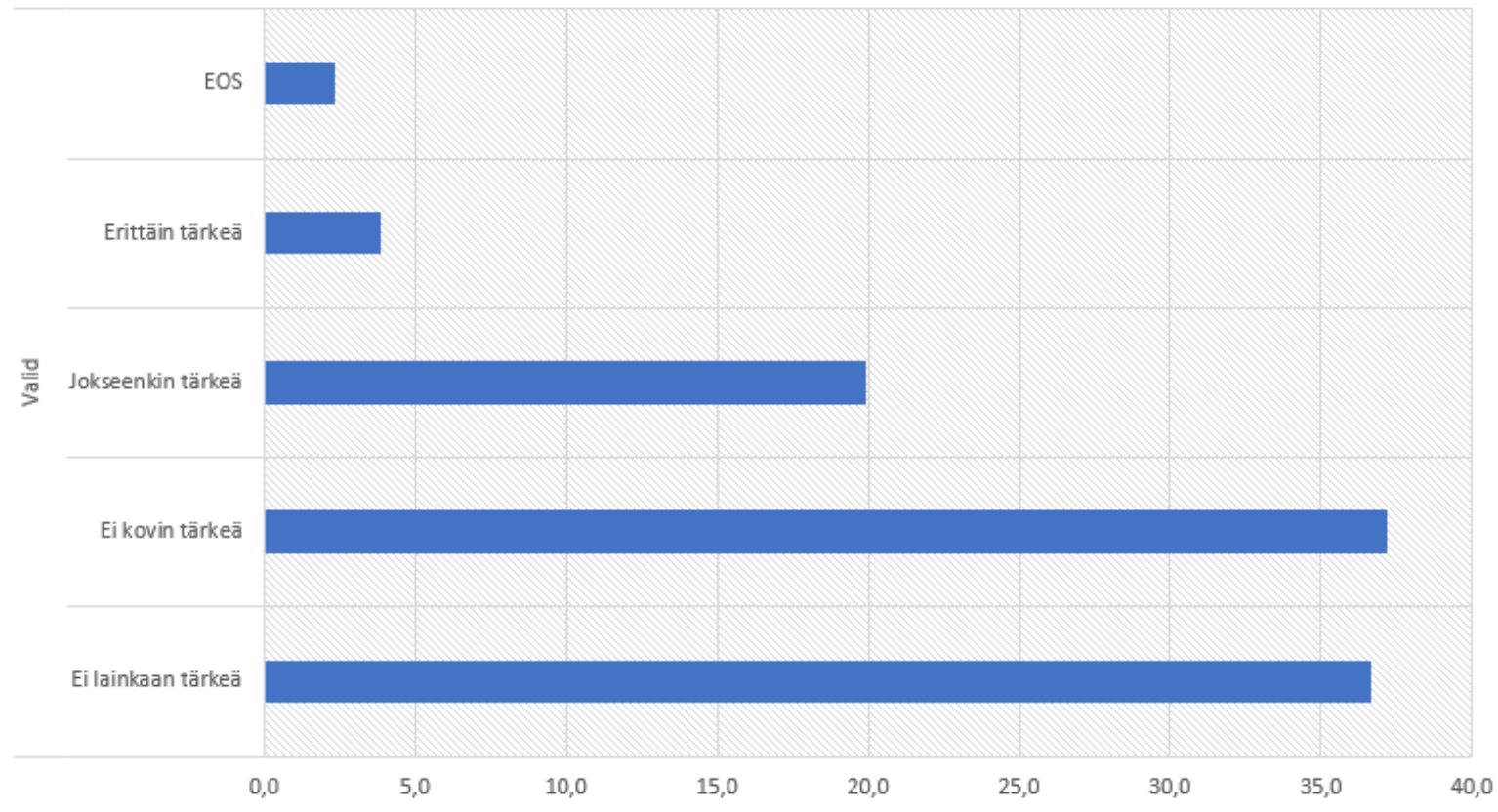
# Tietoarkisto

- ▶ Tietoarkisto vastaanottaa sähköisiä tutkimusaineistoja, jotka sopivat tutkimukseen, opetukseen ja opiskeluun
  - ▶ Saatavilla 1677 aineistoa  
(17.11.2021)
- ▶ Aineistojen tulee olla teknisesti ja juridisesti arkistointikelpoisia



E2 Tutkimus & Suomen Kulttuurirahasto: Suomalaisten identiteetit, arvot ja asenteet 2017-2018 [sähköinen tietoaaineisto]. Versio 1.0 (2021-05-31). Yhteiskuntatieteellinen tietoaarkisto [jakaja]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD3447>

[K3\_16] Kuinka tärkeinä pidät seuraavia asioita identiteettisi kannalta?  
Sosiaalinen media Valid Percent



# Keskustelua somedatan arkistoinnista

- ▶ Breuer J, Bishop L, Kinder-Kurlanda K. The practical and ethical challenges in acquiring and sharing digital trace data: Negotiating public-private partnerships. *New Media & Society*. 2020;22(11):2058-2080. doi:[10.1177/1461444820924622](https://doi.org/10.1177/1461444820924622)
- ▶ Kinder-Kurlanda K, Weller K, Zenk-Möltgen W, Pfeffer J, Morstatter F. Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. *Big Data & Society*. December 2017. doi:[10.1177/2053951717736336](https://doi.org/10.1177/2053951717736336)
- ▶ Fält, Katja: Someaineistojen arkistointi ja jatkokäyttö kaatuvat useimpien alustojen käyttöehtoihin. Tietoarkistolehti, Numero 45 (2/2016)  
[https://www.fsd.tuni.fi/lehti/fi/45/some\\_data.html](https://www.fsd.tuni.fi/lehti/fi/45/some_data.html)

# Twitter-dataa ovat arkistoinneet ainakin GESIS ja UK Data Service

993098201233346562  
993112440492654593  
993112985613807617  
993114241082290176  
993116394089787398  
993118143458545666  
993122103800025089  
993127011081170945  
993129511737856001  
993130546766151680  
993131811147575296  
993132829012787200  
993134287154286592  
993138822316789762  
993139304288440320  
993144056413745153  
993144656920760320  
993146962370269184  
993154177235812354  
993154257208643584  
993161540659568642  
993170217521082370  
993188442019434497  
993194778673209344  
993195955427135488  
993196607972741123  
993201892363784192  
993203769109368832  
993203998936223744  
993212491034447877  
993213107060334594  
993215171760881664  
993216862161948673  
993217650221252610  
993220000000000000

outrageous  
unconscionable  
reprehensible  
absurd  
inexcusable  
preposterous  
ludicrous  
laughable  
disgraceful  
idiotic  
disgusting  
appalling  
despicable  
outlandish  
grotesque  
utterly ridiculous  
scandalous  
unjustifiable  
irresponsible  
ridiculous

# Ongelmat somedatan arkistoinnille

- ▶ Eettiset syyt
- ▶ Tietosuoja
- ▶ Tekijänoikeudet
- ▶ Palvelujen/alustojen **käyttöehdot**

# Etiikka ja tietosuoja

- ▶ Kaikki mikä on julkisesti saatavilla ei välttämättä ole suoraan käytettävissä muihin tarkoituksiin
  - ▶ Yksityisyys- ja julkisuusoletukset erilaisia
    - ▶ Edellyttääkö sisältöihin pääseminen käyttäjäksi rekisteröitymistä vai ei?
    - ▶ Miten käyttäjät itse mieltävät yksityisyyden/julkisuuden? Älä ajattele oman ajatusmaailmasi kautta.
  - ▶ Käyttäjät mieltävät usein sosiaalisen median sisällöt kontekstiinsa sidotuiksi
- ▶ Onko sisällöissä mukana käyttäjien henkilötietoja vai ei?
  - ▶ Tietosuoja-asetusta tulee noudattaa aina, kun käsittelee henkilötietoja



# Tekijänoikeudet

- ▶ Tekijänoikeus suojaa kaikkea tekijänsä omaperäistä ilmaisumuotoa
  - ▶ Valokuvat ja videot saavat aina suojaa
    - ▶ Anekdootti: MAO:125/21 [mao12521.html](http://mao12521.html) (asiaa käsitellään journalismin näkökulmasta, mutta mielenkiintoinen myös tutkimuksen näkökulmasta)
  - ▶ Joskus lyhytkin kirjallinen ilmaisu voi saada teossuojaa
- ▶ Lähtökohtaisesti oikeudenhaltijan lupa edellytetään teosten kappaleiden valmistamiseen ja niiden saataville saattamiseen tai muuhun jakeluun (esimerkiksi arkistointiin osana tutkimusaineistoa)

# Käyttöehdot

- ▶ Määrittelevät käyttäjän ja palveluntarjoajan välistä suhdetta
  - ▶ Esim. miten palvelua ja sen tietoja saa käyttää
- ▶ **Muuttuvat usein**, monipolvisia ja vaikeaselkoisia
- ▶ Palvelun käyttöä ohjaava dokumentaatio voi muodostua useasta eri dokumentista (esim. käyttöehdot, käyttösäännöt, tietosuoja, turvallisuus, opastus jne.)
  - ▶ Merkitys määräytyy tapauskohtaisesti

# Sanoma.fi

- ▶ Helsingin Sanomien artikkelien kommentointiohjeista:
  - ▶ *”Voinko tehdä tutkimusta HS:n keskusteluista? Tietojen kerääminen kommentointiketjuista tutkimuksia ja opinnäytteitä varten on sallittua. Emme kuitenkaan julkaise yhteydenottopyyntöjä tai linkkejä muualla tehtäviin tutkimuksiin tai kyselyihin.”*
- ▶ Huomioitava kuitenkin lainsäädäntö (erityisesti tietosuoja) ja etiikka

# Suomi24

- ▶ Käyttöehtojen tulkinta maallikolle vaikeaa:
- ▶ Suomi24:ä on käytettävä näiden Käyttöehtojen ja Sääntöjen mukaisesti sekä aina siten, että käytöstä ei aiheudu häiriötä muille käyttäjille, Suomi24:lle, Palveluntarjoajille tai kolmansille osapuolille. Suomi24:ä **ei saa käyttää lain tai hyvän tavan vastaisesti**. Käyttäjä sitoutuu käyttämään Suomi24:ä vain näissä Käyttöehdoissa ja Säännöissä nimenomaisesti sallittuihin tarkoituksiin. Mikäli Osapuolten välillä on erimielisyys Suomi24:n toteutuneen tai aiotun käytön tarkoituksesta, määrittää Suomi24 käytön luonteen omassa harkinnassaan.
- ▶ **Käyttäjä saa käyttää Suomi24:ä vain omaan henkilökohtaiseen käyttöönsä.** Käyttäjä ei saa ilman Suomi24:n etukäteen antamaa kirjallista lupaa käyttää Suomi24:ä tai Suomi24:n sisältöä Suomi24:n kanssa samanlaisen palvelun kehittämiseen, tuottamiseen tai markkinoille saattamiseen, jälleenmyydä Suomi24:n sisältöä kolmannelle osapuolelle tai käyttää Suomi24:ä tai sen sisältöä mihinkään kaupalliseen tarkoitukseen. Lisäksi Käyttäjä sitoutuu olemaan ottamatta yhteyttä kehenkään Suomi24:n käyttäjään kaupallisessa tai muussa elinkeinon harjoittamisen tarkoituksessa.
- ▶ Käyttäjällä ei ole oikeutta käyttää Suomi24:ä tai Suomi24:n sisältöä minkään oman palvelunsa osana ilman Suomi24:n etukäteen antamaa kirjallista suostumusta. **Käyttäjällä ei myöskään ole oikeutta kopioida, siirtää, luovuttaa tai saattaa yleisön saataviin, muutettuna tai muuttamattomana, Suomi24:ä tai Suomi24:n sisältöä muutoin kuin tämän Sopimuksen tai lain niin salliessa.** Käyttäjällä ei ole oikeutta ilman Suomi24:n etukäteistä kirjallista suostumusta siirtää Suomi24:n tai Suomi24:n sisällön tähän Sopimukseen perustuvaa käyttöoikeutta kolmannelle osapuolelle.

# Somedatan käyttö edellyttää usein sopimuksia palvelutuottajien kanssa

- ▶ Tiedeyhteisö yrittää varmistaa tutkijoiden pääsyä somedataan sopimuksin:
  - ▶ Social Science One (Harvard University) esim. Facebook aineistojen käyttöön <https://socialscience.one/>
  - ▶ Kielipankki Suomi24 aineistoille <https://www.kielipankki.fi/>
- ▶ Tutkija voi myös yrittää sopia tutkimuskäytöstä suoraan palvelun tarjoajan kanssa esimerkiksi sähköpostitse
  - ▶ Suositeltavaa aina, kun ei pysty tulkitsemaan itse palvelun käyttöehtojen soveltuvuutta tutkimustarkoitukseen

# Jodel.com

## Press/Media/Academic inquiries



فهمني

8 days ago · Updated

Follow

If you have a press/media/academic inquiry please refer to our PR team who will happily assist you with your queries and prospective projects.

All PR requests should be only addressed to: [pr@jodel.com](mailto:pr@jodel.com)

# Suomi24 aineistot Kielipankissa

- ▶ Aller Media Oy ja Kielipankki tehneet sopimuksen, jonka mukaisesti suomi24 aineistot alkaen vuodesta 2001 ovat olleet käytettävissä Kielipankin kautta
- ▶ Uuden omistajan, City Digital, kanssa Kielipankilla vastaava sopimus
  - ▶ Kielipankin Suomi24 aineistojen lisenssien ehtoja noudatettava
  - ▶ Ks. Kielipankin youtube-kanavalla Demos esimerkit Suomi24 aineiston käytöstä

# Perussäännöt

- ▶ Noudata aina
  - ▶ Käyttöehtoja
  - ▶ Lakeja (erityisesti tietosuoja ja tekijänoikeudet)
  - ▶ Ihmistieteiden eettisiä periaatteita



# Aineiston sisäänoton kriteereitä Tietoarkistossa (1)

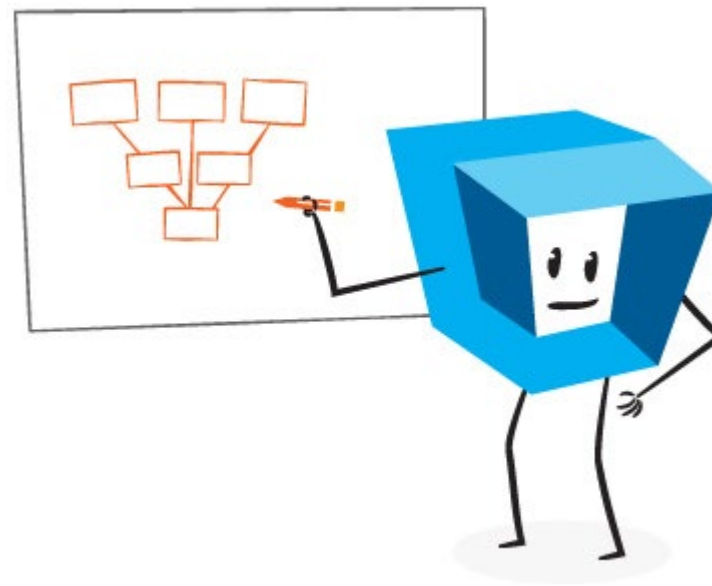
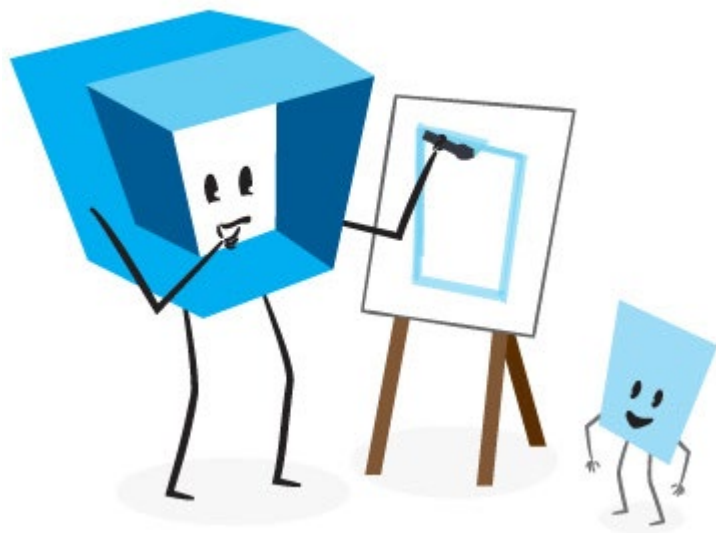
- ▶ Aineiston omistus- ja hallintasuhteiden tulee olla selvillä
  - ▶ Somepalvelun käyttöehtojen tulee mahdollistaa tietojen siirtäminen toiseen palveluun (Tietoarkistoon) edelleen luovutettavaksi tai tutkijan tulee sopia arkistoinnista erikseen palveluntuottajan kanssa
- ▶ Teossuojaa saavien aineistojen arkistointiin on oikeudenhaltijoiden lupa
  - ▶ Poikkeuksena Kopiostosopimuksen alaan kuuluvat aineistot
    - ▶ Tutkimuksessa analysoidut lehtiartikkelit, lehtikuvat, kuvitukset, sarjakuvat
    - ▶ Toimitetut blogit ja joidenkin mediasisältöjen kommenttiketjut, jotka toimitus käyttöehdoissaan määritellyt journalistiseksi sisällöksi (esim. HS ja IS)

## Aineiston sisäänoton kriteereitä Tietoarkistossa (2)

- ▶ 25.5.2018 jälkeen kerätyn alun alkaen **henkilötietoja sisältäneen** anonymisoitavaksi tarkoitetun **aineiston** tietosuojailmoitukseen on sisältynyt maininta aineiston siirrosta Tietoarkistoon tutkimuksen päätyttyä
- ▶ Tarkoittaa käytännössä:
  - ▶ Tutkittavia on informoitu asianmukaisesti henkilötietojen käsittelystä ja aineiston arkistoinnista
  - ▶ Ihmistieteiden eettisiä periaatteita on noudatettu (esim. vapaaehtoinen osallistuminen)
  - ▶ Aineisto tulee olla anonymisoitavissa

# Kiitos!

<https://www.fsd.tuni.fi/>



Tutustu Aineistonhallinnan käsikirjaan!