

ABSTRAKTI

Mari Kleemola

Sähköisen tutkimusaineiston luotettava säilyttäminen : esimerkkinä Yhteiskuntatieteellinen tietoaarkisto

Mari Kleemola, Tampereen yliopisto, mari.kleemola@uta.fi

Hedstromin (1998, 190) mukaan sähköinen säilyttäminen on sellainen yhdistelmä suunnittelua, resurssien allokointia, menetelmiä ja teknologioita, joka varmistaa jatkuvaa arvoa omaavan digitaalisen informaation säilymisen saatavilla ja käyttökelpoisena. Sähköisen säilyttämisen standardi OAIS ja sen suomalainen versio Viitemalli pitkäaikaissäilytysarkistolle (SFS5972) puolestaan korostavat säilyttämisen aktiivista luonnetta. Esimerkiksi "pitkäaikaisuus" tarkoittaa niissä niin pitkää ajanjaksoa, että on otettava huomioon teknologian ja käyttäjäkunnan muutosten vaikutukset (Viitemalli... 2009, 9; OAIS 2009, 1-12; Giaretta 2011, 14). Näissä määrittelyissä esiin tuodut vaatimukset kuvaavat hyvin sähköisten aineistojen pysyvän säilyttämisen monisyistä problematiikkaa.

Tässä tutkimuksessa sähköisten tutkimusaineistojen pitkäaikaissäilytystä tarkastellaan yhteiskunta- ja sosiaalitieteiden näkökulmasta, mutta tulokset ovat yleistettävissä koskemaan muitakin tieteenaloja tai sähköisiä aineistoja yleensä. Eri maiden yhteiskuntatieteelliset data-arkistot ovat sähköisen säilyttämisen edelläkävijöitä (Doorn & Tjalsma 2007) . Niiden synty ajoittuu viime vuosisadan puoliväliin, jolloin tietokoneiden kehitys loi pohjan yhteiskuntatieteilijöiden suosimalle survey-tutkimukselle. Laajojen survey-aineistojen keruu ja erilaiset aineistojen käyttötavat loivat tarpeen säilyttää ja hallita aineistoja, mikä puolestaan johti ensimmäisten data-arkistojen perustamiseen 1950- ja 1960-luvuilla. Nykyään data-arkistot elävät jatkuvan muutoksen aikaa muun muassa nopeasti muuttuvan tietoteknologian vuoksi.

Keskustelua yhteiskuntatieteellisten tutkimusaineistojen arkistoinnista ja jatkokäytöstä on käyty monissa julkaisuissa (ks. esim. Hyman 1972; Sharing Research Data 1985). Aineistoilla ajatellaan olevan jatkuvaa arvoa, ja aineistojen arkistointi on tärkeää muun muassa tutkimuksen validiteetin ja reliabiliteetin kannalta. Yhteiskuntatieteissä tutkittavat tapahtumat ja asiat eivät tavallisesti ole toistettavissa, joten aineistoa ei voi kerätä uudelleen tai koetta toistaa täysin samanlaisena. Hyvä esimerkki tästä on vaalitutkimus. Vaalit ovat toistuva tapahtuma, mutta yhteiskunta ja asenneilmasto yleensä muuttuvat vaalien välillä, eikä yhden vaalin aikana kerätyllä tutkimusaineistolla voi suoraviivaisesti todentaa sellaisia tutkimustuloksia, jotka perustuvat vaikkapa edellisten vaalien aikana kerättyyn dataan. Alkuperäisen tutkimusaineiston säilyttäminen saatavana ja käyttökuntoisena onkin oleellinen osa tutkimusprosessia.

Tutkijoilla ei voi olettaa olevan resursseja sähköisten tutkimusaineistojen pitkäaikaissäilytykseen. Antamalla aineistonsa data-arkistolle tai tallennuspalveluun tutkija voi keskittyä varsinaiseen työhönsä eli tutkimukseen. Tutkijan tai tutkimusorganisaation näkökulmasta kysymykseksi nouseekin, miten voi tietää, onko säilytyspalvelun tarjoaja luotettava. Onko sillä tarvittavat valmiudet aineiston pitkäaikaiseen säilyttämiseen niin, että aineisto on tulevien tutkijasukupolvienkin käytettävissä? Toisaalta voidaan kysyä, miten data-arkisto tai tallennuspalvelu voi osoittaa olevansa luotettava. Miten se voi varmistaa, että sen menettelytavat, toiminta ja prosessit ovat ajan- ja asianmukaiset ja että sen huomaan uskottu tutkimusaineisto säilyy?

Tässä tutkimuksessa kartoitetaan keinoja ja välineitä, joilla voidaan arvioida data-arkiston tai tallennuspalvelun luotettavuutta sähköisen aineiston säilyttäjänä. Erityisen kiinnostuksen kohteena ovat standardit ja muut laajasti hyväksytyt mallit, muun muassa edellä mainitut OAIS-viitemalli (ISO14721:2003 Open Archival Information System Reference Model) ja sen suomalainen versio (SFS5972) sekä ISO:n tämän vuoden alussa julkaisema luotettavan sähköisen arkiston vaatimukset määrittävä standardi, nk. TDR-tarkistuslista (ISO 16363:2012 Audit and certification of trustworthy digital repositories).

Esimerkkitapauksena on Yhteiskuntatieteellisen tietoarkiston luotettavuuden arviointi. Tietoarkisto on valtakunnallinen palveluinfrastrukturi, joka arkistoi ja välittää sähköisiä tutkimusaineistoja tutkimukseen, opetukseen ja opiskeluun. Suomen Akatemian rahoittamassa Tietoarkisto 2015 –projektissa on verrattu tietoarkiston toimintoja OAIS-viitemalliin ja tarkasteltu, kuinka hyvin tietoarkisto täyttää ISO 16363 -standardin vaatimukset. Esityksessäni kuvailen arviointiprosessia ja sen tuloksia ja vertaan niitä muiden vastaavien organisaatioiden tuloksiin (muun muassa Vardigan & Whiteman 2007, Downs & Chen 2012, Lyle 2012). Lopuksi pohdin arviointivälineitä ja niiden hyödyntämisen etuja ja haittapuolia.

Lähteet

Doorn, Peter & Tjalsma, Heiko (2007). Introduction: archiving research data. – *Archival Science* 7(1):1-20. DOI:10.1007/s10502-007-9054-6

Downs, Robert R. & Chen, Robert S. (2012). Improving the Trustworthiness of an Interdisciplinary Scientific Data Archive. A presentation at the 38th Annual IASSIST Conference, Washington DC, June 06, 2012.

Giaretta, David (2011). *Advanced Digital Preservation*. Heidelberg: Springer-Verlag Berlin.

Hedstrom, Margaret (1998). Digital Preservation: A Time Bomb for Digital Libraries. – *Computers and the Humanities* 31(3):189-202. DOI 10.1023/A:1000676723815

Hyman, Herbert (1972). *Secondary Analysis of Sample Surveys: Principles, Procedures, and Potentialities*. New York: John Wiley & Sons.

Lyle, Jared (2012). TRAC Self-Assessment at ICPSR. A presentation at the 38th Annual IASSIST Conference, Washington DC, June 06, 2012.

OAIS Pink Book. Reference Model for an Open Archival Information System (OAIS) Draft Recommended Standard. CCSDS 650.0-P-1.1 (Pink Book), Issue 1.1 August 2009. URL: <http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf> (20.2.2012)

Sharing Research Data (1985). Toim. Fienberg, Stephen E., Martin, Margaret R. & Straf, Miron L. Committee on National Statistics Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, D.C.: National Academy Press.

TDR Magenta Book. Audit and Certification of Trustworthy Digital Repositories. CCSDS 652.0-M-1 (Magenta Book), Issue 1 September 2011. URL:
<http://public.ccsds.org/publications/archive/652x0m1.pdf> (10.11.2011)

Vardigan, Mary & Whiteman, Cole (2007). ICPSR meets OAIS: applying the OAIS reference model to the social science archive context. – *Archival Science* 7(1):73-87. DOI:
10.1007/s10502-006-9037-z

Viitemalli pitkäaikaissäilytysarkistolle. OAIS-standardin suomalainen versio (2009). SFS5972. Helsinki : Suomen standardisoimisliitto, 14.12.2009.