

Semanttinen web, ontologiat ja avoin tieto – talkoot avoimen yhteiskunnan puolesta

Johanna Enqvist

“Tieto lisää tuskaa”, totesi roomalainen reetori Quintilianus ajanlaskumme ensimmäisellä vuosisadalla. ”Tieto on valtaa”, julisti puolestaan brittifilosofi ja valtio-opinut, modernin tieteen isänä pidetty Francis Bacon noin 400 vuotta sitten. 2000-luvulla tieto ja sen vapaa saatavuus ovat itseisarvoja, mutta myös tietotulvaa ja irrallisena vyöryvää dataa. Meidän aikamme haasteita ovat analogisen tiedon muuntaminen digitaaliseksi ja digitaalisen tiedon hallinta niin, että se ei hallitse meitä. Demokratiassa valta sekä oikeus tietoon on kaikilla kansalaisilla. Avoin tiedonkulku ei kuitenkaan toteudu itsestään, vaan vaatii aktiivista poliittista tahtoa ja valtiiovallan tukea, asiantuntijoiden ja organisaatioiden yhteistyötä sekä toimivan teknologian ja infrastruktuurin. Suomessa on vahvoja yhteis- ja ilmaistoinnin perinteitä, kuten talkookulttuuri, jokamiehenoikeus ja kirjastolaitos. Tässä samassa hengessä tulisi suhtautua tietoon: avoimena, ilmaisena ja kaikille ihmisille kuuluvana perusoikeutena. Tieteentekijät ovat tiedon tuottajina ja ennen kaikkea yhteiskunnan palvelijoina merkittävässä roolissa tiedon avaamiseen tähtävissä talkoissa.

Tietoyhteiskunnan rakennuspalikoita

Viestintä on yksi inhimillisen toiminnan peruselementeistä, tiedon ja sanomien siirtoa, yhteisöllisyyttä ja yhteisen sosiaalisen todellisuuden

rakentamista. Joukkoviestinnän käsite on 2000-luvulla osittain korvautunut käsitteellä media. Tämä heijastaa muutostavassa, jolla viestintä ymmärretään. Viestintä on enemmän kuin tiedon siirtoa ja media enemmän kuin tiedon siirtämisen väline: media on myös sisältöjen merkityksiä sekä merkitysten jakamista, välittämistä ja tulkittaa kokonaisuudessaan. Mediaan kuuluvat myös sen toimintaan ja käyttöön liittyvät sosiaaliset ja kulttuuriset tavat. Sanastokeskuksen määritelmän mukaan sosiaalinen media on ”tietoverkkoja ja tietotekniikkaa hyödyntävä viestinnän muoto, jossa käsitellään vuorovaikutteisesti ja käyttäjälähtöisesti tuotettua sisältöä ja luodaan ja ylläpidetään ihmisten välisiä suhteita” (TSK Sanastokeskus 2010). Termi on sikäli epäonnistunut, että sosiaalisuus kuuluu myös vanhoihin medioihin. Merkityksellisempää sosiaalisen median nousussa on reaaliaikaisuus sekä riippumattomuus ajasta ja paikasta. Sosiaalisen median suosiossa on joka tapauksessa kyse pelkkää teknologista kehitystä syvällisemmästä muutoksesta yhteiskunnallisessa ilmapiirissä, asenteissa sekä viestintään liitetyissä tarpeissa ja odotuksissa. Sosiaalisen median sovelluksia on tuhansia, joista tunnetuimpia ovat tällä hetkellä verkostopalvelu Facebook, mikroblogipalvelu Twitter, videoiden jakamiseen tarkoitettu YouTube ja yhteistuotantona luotu tietosanakirja Wikipedia.

Viestinnän suhde demokratiaan ja julkisuteen on yksi oleellisimpia kysymyksiä vies-

tinnän tutkimuksessa. Media on keskeinen toimija myös demokratian toteuttamisessa. Demokratia voidaan tässä yhteydessä määritellä laajaksi elämäntavan periaatteeksi, jonka ihanteisiin kuuluvat mm. sananvapaus, kansalaisjournalismi, joukkoilyn hyödyntäminen sekä avoin, vapaasti saatavilla oleva tieto. Tietoyhteiskunta (myös informaatioyhteiskunta, verkkoyhteiskunta) on näille ihanteille pohjautuva jälkiteollinen ja postmoderni yhteiskunta, joka vertautuu käsitteenä maatalous-, teollisuus- ja palveluyhteiskuntaan. Tietoyhteiskunnan perustana on uusi teknologia ja siihen liittyvät toimintatavat; informaation tuottaminen, jakelu ja käsittely ovat läsnä kaikissa yhteiskunnan toiminnoissa. Liikenne- ja viestintäministeriön vetämä Arjen tietoyhteiskunnan neuvottelukunta on kansallisten tietoyhteiskuntastrategioiden (1995, 1998, 2007, 2010) valmistelija ja toteuttaja Suomessa. Tuoreimmassa strategiassa Tuottava ja uudistuva Suomi – Digitaalinen agenda vuosille 2011–2020 esitetään keskeiset tavoitteet tietoyhteiskunnalle sekä keinot niiden saavuttamiseksi (Arjen tietoyhteiskunnan neuvottelukunta 2010).

Data jalostuu tiedoksi

Platonin Theaitetos-dialogiin (Plato 2007) perustuvan tietoteoreettisen määritelmän mukaan tieto on hyvin perusteltu tosi käsitys (Wikipedia). Tämä käsitys tiedon olemuksesta on edelleen pätevä joissakin yhteyksissä, vaikka tieto on kaikkien käsitteiden tavoin kontekstisidonnainen ja ajassa muuttuva. Suomen kielessä termejä data, informaatio (engl. information) ja tieto (engl. knowledge) käytetään usein synonyymeina. Informaatio-tutkimuksessa data tarkoittaa merkeistä ja symboleista koostuvaa, (digitaalisessa muodossa olevaa) raaka-ainetta, potentiaalista informaatiota. Informaatio on puolestaan datasta jalostettua merkityksellisempää sisältöä, joka voidaan välittää. Tieto taas on informaatiota, jonka vastaanottaja on tulkinnut ja jolle antanut merkityksen. Tiedon arvoketju: Data

à Informaatio à Tieto à Tietämys à Viisaus kuvaa prosessia, jossa sinällään merkityksetön data ja irrallinen informaatio rikastuvat ihmistä hyödyttäväksi tiedoksi ja lopulta viisaudeksi. (Poikola ym. 2010, 13–14.) Avoimesta tiedosta (engl. Open Data) puhuttaessa tarkoitetaan yleensä avointa dataa, raaka-ainetta joka on kaikkien jalostettavissa. Datan arvoa voidaan siis kasvattaa sitä jakamalla.

Avoimien tietojen tuottaminen ja jakaminen on mitä tahansa digitaalisia dokumentteja, tietokantoja, sopimuksia tai audiovisuaalisia tallenteita sekä ei-elektronisessa muodossa olevien sisältöjen digitointia. Avoimeen tietoon eivät sisälly esimerkiksi yksilönsuojan piiriin kuuluvat tai valtion kannalta arkaluonteiset tiedot. Käsitteeseen kuuluvat tiedon avoimeen tuottamiseen ja jakeluun liittyvät periaatteet ja menetelmät, mutta se on myös lähestymistapa ja asenne. Avoimen tiedon kriteerit ovat: löydettävyys, kokonaisuus, käyttöehtojen tasa-arvoisuus, alkuperäisyys ja ajantasaisuus, laillinen ja vapaa uudelleenkäytettävyys, maksuttomuus, koneluettavuus, formaatin avoimuus ja ymmärrettävyys. (Poikola ym. 2010, 34–38). Avoin tieto ja sosiaalinen media tukevat toisiaan ja voivat oikein hyödynnettyinä lisätä minkä tahansa organisaation ja sen tietovarantojen vaikuttavuutta. Avoimen datan ”ekosysteemissä” raakadatalle ei käydä kauppa, joten tiedon hankkimisesta vapautuneet resurssit voidaan kohdistaa sen hyödyntämiseen. Avoin tieto voi tuottaa uusia palveluita, tutkimusta ja tietoa, joista osalla on kaupallista arvoa, osalla yhteiskunnallista ja sivistyksellistä. Hallinnon organisaatiot, kansalaiset ja yritykset ovat sekä tiedon käyttäjiä että tuottajia; hallinnon rooli on tuottaa infrastruktuuria kaikkien käytettäväksi. Perinteisten liiketoimintamallien ja rahatalouden ulkopuolisen tuotannon merkitys kasvaa jatkuvasti: ihmiset, jotka tarvitsevat palvelun, voivat tulevaisuudessa itse ideoida, luonnostella ja toteuttaa sen. (Poikola ym. 2010, 15, 25, 84.)

Välineenä tieto- ja viestintäteknologia

Tietoyhteiskunnan teknologinen perusta Internet on maailmanlaajuinen tietoverkko, joka yhdistää paikallisia tietoverkkoja toisiinsa. Pienellä alkukirjaimella kirjoitettuna internet on yksi viestintävälineistä, kuten televisio tai radio. Internet ei ole kuitenkaan media samassa mielessä kuin vanhat viestimet, vaan eräänlainen metamedia, joka voi ilmaista kaiken digitaaliseen muotoon muunnettavan. Vuonna 1990 julkistettu, Tim Berners-Leen luoma World Wide Web (WWW/Web) on Internetissä toimiva linkkeihin perustuva hypertextijärjestelmä, yksi monista Internetin palvelumuodoista. Hypertekstiä luetaan selaimella, joka hakee verkkosivuiksi kutsuttuja HTML (Hypertext Markup Language) -kielellä koodattuja dokumentteja verkkopalvelimilta ja esittää ne ihmisille luettavassa muodossa. Verkkovälitteisen viestinnän vallankumouksellisuus pohjautuu osaltaan juuri lineaarisen tekstin korvautumiseen hypertekstillä. Hyperteksti on luettavissa lukijan päättämässä järjestyksessä, linkkien kautta on mahdollista saavuttaa lähes rajattomasti uusia tekstejä tai palata edellisiin.

Verkon toimintaa kehittää ja ohjaa vuonna 1994 perustettu kansainvälinen W3C-konsortio, jonka johdossa on WWW:n isä Tim Berners-Lee (World Wide Web Consortium). Berners-Leen semanttiseksi webiksi kutsumaa kehityslinjaa ryhdyttiin tutkimaan, kun tiedonhaku miljardien dokumenttien massasta sekä tiedonhallintajärjestelmien yhteensovittaminen alkoi tuottaa hankaluuksia. Verkkosivujen lisäksi voidaan linkittää dataa ja sisältöjä (Linked Data), mikä on seuraava askel tietoverkkojen kehityksessä. Käsitteillä semanttinen web ja Web 2.0 viitataan molemmilla nykyistä WWW:a täydentävään sisältölliseen metatietokerrokseen, joka kertoo koneille verkossa olevien sisältöjen merkityksen. Nykyinen verkko (Web of Pages) linkittää sivuja ihmiselle, joka tekee hakuja perinteisellä selaimella. Linkitetyn datan

verkko (Web of Data) linkittää tietoa koneelle ja ihmiselle siten, että hakuja voidaan tehdä myös semanttisella sovelluksella. (Hyvönen 2011.) Linkitetyn datan verkko on suunniteltu kasvamaan ja kehittymään orgaanisesti, kun ihmiset ja organisaatiot spontaanisti lisäävät omia resurssejaan verkkoon ja linkittävät niitä toisiinsa. Tämä edellyttää myös, että toimijat luovat sanastoja kuvaamaan sisältöjä ja liittävät dokumentteihin metatietoa sisältöistä. (Poikola ym. 2010, 73.) Linkitetty data mahdollistaa sisältöjen aiempaa tehokkaamman linkittämisen, jakamisen ja älykkäämpien verkkosovellusten toteuttamisen. Edellytyksenä on kuitenkin linkitettävän datan saattaminen koneluettavaan muotoon.

Semanttisen webin ontologiat ovat työkalu koneluettavan tiedon tuottamisessa, digitaalisen ajan versio jo pitkään käytössä olleista asiasanastoista ja tesaursuksista. Siinä missä sanastot listaavat ja luokittelevat termejä, ontologian perusyksikkö on käsite. Tarkoitteet ovat konkreettisia (esim. esineet) tai abstrakteja (esim. yhteiskunta) todellisuuden ilmiöitä, joita ajatellessamme valitsemme joukon ominaisuuksia. Nämä ominaisuudet eli käsitepiirteet yhdistyvät mielikuvaksi eli käsitteeksi. Käsitteet ovat siis abstrakteja yksilöllisiä mielikuvia, joista puhumiseen tai kirjoittamiseen tarvitaan määritelmiä ja termejä. Määritelmät ovat käsitteen kuvauksia ja termit niihin viittaavia nimityksiä. (Suonuuti 2006, 11.) Filosofian alalla ontologia on yksinkertaistetusti 'oppi olevaisesta' (Wikipedia) ja semanttisen webin ontologian voi analogisesti ymmärtää käsitteelliseksi ontologisoidun aihealueen olemuksesta.

Ontologia on valitun aihepiirin käsitteistöä ja käsitteiden välisiä suhteita esittävä täsmällinen kuvaus. Käytännössä ontologia muodostetaan yleensä tiettyä tarkoituspäätä, esimerkiksi verkkosovellusta varten. Ontologia mahdollistaa tietojen sisältöllisen yhdistelyn ja tietokoneella tehtävän koneellisen päättelyn – tiedon indeksoinnin, systematisoinnin ja harmonisoinnin. Ontologian avulla

voidaan siis kuvata tietosisältöjä (annotointi) joko niin, että käsitteet poimitaan aineiston metatietoihin ns. käsin tai oppivien ohjelmien avulla. Ontologia ei ole koskaan lopullinen esitys, vaan dynaaminen konstruktio, joka vaatii ylläpitoa. Ontologioiden muodostamista ja esittämistä varten on kehitetty metakieliä ja -standardeja sekä ontologiatekniikoita. Kulttuuriperintöön liittyy erityisesti ISO-standardin statuksen saanut CIDOC–CRM-malli (CIDOC CRM – International Committee for Documentation Conceptual Reference Model). Ontologioiden muodostamisessa yleisesti käytetty ohjelma, Stanfordin yliopistossa kehitetty Protégé-ontologiaeditori, on ladattavissa ilmaiseksi verkossa.

Ajan henki muuttuu poliittiseksi tahdoksi

Tuhansia salaisia asiakirjoja verkossa julkaissut WikiLeaks-sivusto on hyvä esimerkki verkottuneen yhteiskunnan ja digitaalisen vallankumouksen synnyttämästä ilmiöstä. Sen toiminta on mahdollista maailmassa, jossa valtavia määriä informaatiota ja dokumentteja voidaan siirtää, säilöä ja tutkia sähköisessä muodossa. Teknisten ratkaisujen ja välineiden lisäksi WikiLeaksin toiminta pohjautuu ideologiaan, jonka mukaan kaiken yhteiskunnallisesti ja poliittisesti merkittävän tiedon tulee olla kaikkien saatavilla. Ihmisillä on oikeus tietää, ja tiedon läpinäkyvyys on avain parempaan yhteiskuntaan. Sosiaaliseen mediaan liitetyt arvot – demokraattisuus, vuorovaikutus, vapaaehtoisuus, avoimuus, tasa-arvoisuus ja ilmaisuus – katsotaan yhä tärkeämmiksi kaikissa yhteiskunnan toiminnissa. Näitä hyveitä voidaan vaalia kaikilla tasoilla yksittäisten kansalaisten toiminnasta valtionhallinnon organisaatioiden ja suur-yritysten viestintäkulttuuriin.

Valtio on Suomessakin seurannut kehitystä ja tarttunut toimeen. Jyrki Kataisen hallituksen hallitusohjelmaan on kirjattu monia avoimeen tietoon ja sähköiseen viestintään liittyviä periaatteita ja kehitystavoitteita. Edel-

lisen hallituksen maaliskuussa 2011 vahvistama periaatepäätös julkishallinnon digitaalisten tietoaineistojen saatavuudesta (Liikenne- ja viestintäministeriö 2011) on viety hallitusohjelmaan erittäin suorasanaisesti: ”Julkiset tietoaineistot saatetaan koneluettavassa muodossa avoimesti saataville ja jatkokäytettäviksi.” (Valtioneuvoston kanslia 2011, 78, 50). Julkinen valta velvoitetaan siis levittämään hallinnoimaansa tietoa aktiivisesti ja maksutta. Muut tiedon tuottajat voivat menetellä omasta aloitteestaan samalla tavalla. Syyskuussa 2011 voimaan tullut tietohallintolaki (634/2011) ei edellytä pelkästään julkishallinnon tuottamien tietojen avointa jakamista, vaan jakamisen tehostamista ja saavutettavuuden parantamista semanttisen yhteismitallistamisen avulla. Valmisteilla on myös ”Julkishallinnon standardisalkku” (JHS), jonka tavoitteena on listata julkisen hallinnon tietohallinnossa käytettäväksi suositellut standardit käyttökohteineen. Tässä yhteydessä voidaan määrätä käytettäväksi standardoituja sanastoja.

OECD:n (Organisation for Economic Cooperation and Development) tiede- ja teknologiaministerit hyväksyivät tammikuussa 2004 Pariisissa julistuksen julkisrahoitteisten sähköisten tutkimusaineistojen käytön avoimuudesta (Declaration on Access to Research Data from Public Funding). Julistus tarkoittaa poliittista sitoutumista julkisrahoitteisten sähköisten tutkimusaineistojen avoimen saatavuuden kehittämiseen. Julistuksesta ja siihen liittyvistä suosituksista (OECD 2007) kerättiin mielipiteitä kansallisilla kuulemiskierroksilla. Suomalaisten kulttuurin ja yhteiskunnan tutkimusta edustavien tutkijoiden, yliopisto-organisaatioiden sekä tutkimusrahoittajien näkemyksiä on esitelty tuoreeltaan vuonna 2007 ilmestyneessä raportissa Julkisrahoitteisen tutkimusdatan avoin saatavuus ja elinkaari (Borg ja Kuula 2007). Suomessa tutkimuksen tietoaineistoja koskevaa selvitystyötä on ohjannut opetus- ja kulttuuriministeriön asettama Tutkimuksen tietoaineistot-selvityshanke. Hankkeen tavoitteena on ol-

lut kokonaiskuva Suomen tilanteesta sekä suunnitelma (tiekartta) tietoaineistojen saatavuuden ja säilytyksen kehittämiseksi. Sekä hallitusohjelmaan viety kirjaus että valtioneuvoston periaatepäätös ovat suunnitelmassa esitettyjä ja jo toteutettuja toimenpide-ehtouksia. (Opetus- ja kulttuuriministeriö 2011.) Tiekartan toimenpide-ehtoukset pohjautuvat puolestaan 2010 julkaistuun taustaselvitykseen Tutkimuksen tietoaineistot – Olennaisen käsikirja päättäjille, jossa listattiin keskeisiä suosituksia tutkimuksen tietoaineistojen saatavuuden ja hyödyntämisen parantamiseksi (CSC – Tieteen tietotekniikan keskus 2010).

Monipolvisista selvityksistä, suosituksista ja ehdotuksista huolimatta Suomessa ei vielä ole varsinaisia tietovarantojen avaamista koordinoivia toimielimiä. Tietoyhteiskunnan kehittämiskeskus TIEKE ry. on luomassa avoimen datan foorumia, joka tulee olemaan ainakin keskustelua ja toimijoita kokoava kohtauspaikka. Selvitettävänä on lukuisia lainsäädännöllisiä, taloudellisia ja teknisiä kysymyksiä liittyen mm. tietoturvaan, liikesalaisuuksiin, tekijänoikeuksiin ja valtionhallinnon IT-arkkitehtuuriin. Tiedon avaamisen ja avoimen tiedon hyvät käytännöt ovat niin ikään vasta muotoutumassa. Lainsäädännön selkeyttäminen on kuitenkin kirjattu hallitusohjelmaan: tarkoituksena on valmistella uusi tietoyhteiskuntakaari, johon kootaan keskeiset sähköistä viestintää ja tietoyhteiskunnan palvelujen tarjontaa koskevat säännökset (Valtioneuvosto 2011, 51). On liian varhaista arvioida mihin julkishallinnon tietovarantojen avaaminen lopulta johtaa. Maailma ja toimintaympäristöt muuttuvat vauhdilla, jota yksikään instanssi ei pysty kattavasti seuraamaan. Tästä syystä yksityiskohtia ei ole välttämättä mahdollista tai järkevääkään suunnitella kovin kauas tulevaisuuteen.

Ensin työ, sitten huvi: hankkeita ja projekteja

”Talkoot ovat työapua tarvitsevan auttamiseksi järjestetty yleinen, vapaaehtoinen

ja palkaton työntekotilaisuus, johon liittyy autettavan tarjoama kestitys ja päätteeksi tanssit.” (Wikipedia)

Tietoyhteiskunnan kehitys on toistaiseksi ollut paljolti teknologista infrastruktuurin kehitystä. Kaupalliset suuryritykset ovat toiminnallaan määritelleet verkkoviestinnän suuntaviivoja ja periaatteita. Taloudelliset näkökohdat ovat olleet määrääviä kaikkialla maailmassa informaatioteknologiaa kehitettäessä ja omaksuttaessa. (mm. Van Dijk 2006.) Valtiot, julkisyhteisöt ja erilaiset tutkimuskonsortiot ovat kuitenkin viime vuosina ryhtyneet aktiivisesti muodostamaan yhtenäisiä visioita, strategioita ja politiikkaa sekä konkreettisia kehityshankkeita. Lokakuussa 2011 alkanut EUDAT-projekti (EUDAT – European Data Infrastructure) on Euroopan tiede- ja tutkimusyhteisöjen yhteistyöprojekti, jonka tavoitteena on tuottaa tutkijoiden tarpeita vastaava tietoinfrastruktuuri (Collaborative Data Infrastructure CDI) hallitsemaan nopeasti lisääntyvää dataa. Projektia koordinoi suomalainen CSC – Tieteen tietotekniikan keskus Oy (Tutkimuksen e-infrastruktuurista Suomessa, ks. CSC 2008).

Semanttisen webin ja ontologioiden osalta tutkimuksen uranuurtajia Suomessa ovat Aalto-yliopiston ja Helsingin yliopiston tietojenkäsittelytieteen laitoksen SeCo-tutkimusryhmä (Semantic Computing Research Group) ja kansallinen FinnONTO-hanke (2003–2012). Hankkeessa on kehitetty ja pilotoitu kansallista semanttisen webin ontologiainfrastruktuuria, jonka osaksi myös kaikki jatkossa luotavat ontologiat voidaan liittää ONKI-ontologiakirjaston kautta. Hankkeen tavoitteiden mukaisesti asiansaastoista siirrytään ontologioihin, kansalliset ontologiapalvelut tuodaan verkkoon ja palvelut perustuvat avoimille standardeille ja sisällöille. Tavoitteiden avulla pyritään edistämään sekä liiketoimintaa että julkisia palveluita. (Hyvönen ym. 2008.) Ontologiatyö aloitettiin käytännössä Yleisen suomalaisen asiansaaston (YSA) ontologisoinnilla (YSO). YSO on ns. yläonto-

logia, jonka avulla erityisalojen tai -aiheiden ontologiat voidaan liittää toisiinsa käyttäjälle näkyväksi yhdeksi kokonaisuudeksi (KOKO). Museoalan asiasanasto (MASA) on jo ontologisoitu (MAO), mutta se on erittäin puutteellinen arkeologisen kulttuuriperinnön ja kulttuuriympäristön osalta. Tiettyä alaa tai aihepiiriä käsittelevien ontologioiden luominen toteuttaa myös Kansallinen digitaalinen kirjasto (KDK) -hankkeen tavoitteita, jotka taas liittyvät Euroopan digitaalisen kulttuurin ja taiteen keskuksen Europeanan toimintaan (Europeana).

Vaikka kulttuuriperintö ja museoala ovat hyvin edustettuina FinnONTO-hankkeen yhteydessä syntyneissä julkaisuissa sekä käyttöliittymien demoversioissa, on lähestymistapa yleisesti ottaen ollut teknisesti painottunut. Poikkeuksena mainittakoon Suvi Kettula (2009), joka kehitti väitöskirjassaan semanttisen webin ontologisen tekstiilikäsitteistön ja kokeili sen liittämistä museoiden luettelointitietoihin. FinnONTO-hankeessa kehitetyt MuseoSuomi ja Kulttuurisampo ovat ontologiapohjaisia semanttisia portaaleja, jotka hyödyntävät yhteistyötahojen aineistoja. Näissä sovelluksissa insinööriä suunnittelun ongelmat ovat ilmeisiä, mutta myös aineiston kuvailutietojen ja luettelointikäytäntöjen yhteismitattomuus ja epäloogisuus huonontavat osaltaan niiden käytettävyyttä. Portaalien tarkoituksena on tosin tässä vaiheessa vasta demonstroida semanttisten sovellusten periaatteita ja mahdollisuuksia. FinnONTO jatkuu Linked Data Finland -hankkeena (2012–2013), jonka tavoitteena on kansallisten tietoaisteistojen julkaiseminen avoimena linkitettyinä tietona (Linked Open Data) sekä avoimen lähdekoodin teknologian kehittäminen. Käytännön pilotointi sekä aineistojen että sovellusten osalta jatkuu myös edelleen. Tutkimuskonsortiossa ovat mukana mm. Kansalliskirjasto, Kotus, Metsähallitus, Museovirasto, Sanastokeskus TSK, SKS ja YLE. (Hyvönen 2011.)

Avoim tiede, avoin arkeologia

Tieteentekijöiden tehtävä on tuottaa ja jakaa tietoa. Jokaisen tieteenharjoittajan ja tutkimusorganisaation on kannettava tämä vastuunsa tiedeyhteisölle ja muulle yhteiskunnalle. Tiedon arvo kasvaa jaettaessa ja mitä nopeammin tieto jaetaan, sitä nopeammin myös tiede etenee. Avoimuus tarkoittaa tiedeessä ennen kaikkea avointa, rajoituksetonta (ja maksutonta) pääsyä tuloksiin; sekä raakadata että tutkimustulokset ovat vapaasti uudelleenkäytettävissä. Avoimuuteen kuuluu myös tutkimusmenetelmien ja -prosessin läpinäkyvyys. ”Avoimen muistikirjan tiedeessä” (open notebook science) (Heikkinen 2007) meneillään olevasta tutkimuksesta voidaan kertoa esimerkiksi blogissa tai wikissä. Tutkimusprosessia voi tällöin seurata ja kommentoida kuka tahansa, aihetta ja tutkimuskysymyksiä voidaan käsitellä avoimissa ja julkisissa verkkokeskusteluissa. Tulevaisuudessa tutkimus on todennäköisesti entistä monitieteisempää ja vaatii eri alojen osaajista koostuvia työryhmiä. Perinteisen sektorijajattelun ylittävä tieteellinen yhteistyö yliopistojen, valtion ja teollisuuden tutkimuslaitosten sekä tieteellisten seurojen ja harrastajien kesken on myös yleistymässä. Nämä vaativat tieteenalakohtaisten ja rahoituksen luomien ”siilojen” purkamista, avoimempaa asennetta ja lähestymistapaa. (Forsström 2011.) Monitieteisyys, yhteistyö ja yhteiset tutkimusintressit esimerkiksi luonnontieteilijöiden kanssa ovat kuuluneet arkeologiseen tutkimukseen jo kauan. Toisaalta yhteishankkeita muiden humanististen tieteenalojen edustajien tai yhteiskuntatieteilijöiden kanssa toteutetaan yllättävän vähän, vaikka arkeologian tutkimuskohteena kuitenkin on ihminen ja yhteiskunta.

Tutkimusprosessi vaatii monista eri lähteistä ja useilla menetelmillä tuotettujen tietojen ja aineistojen yhdistelyä. On arkeologisen kulttuuriperinnön etu, että nämä tiedot olisivat mahdollisimman avoimesti ja kattavasti

käytettävissä tutkimuksen ja suojelun tarpeisiin. Museovirasto vastaa Suomessa arkeologian alan keskeisten tietosisältöjen hallinnoinnista ylläpitämällä valtakunnallista muinaisjäännösrekisteriä ja löytöluetteloa sekä arkistoimalla arkeologisen kulttuuriperinnön tutkimukseen liittyviä tutkimusraportteja. Tutkimusraporttien lisäksi Museovirasto ja sopimuksen tehneet maakuntamuseot tuottavat muinaisjäännösten suojeluun ja tutkimukseen liittyviä viranomaislausuntoja. Sähköisessä muodossa olevat dokumentit: kenttätutkimusraportit, ajoitus- ja analyysitulokset, viranomaislausunnot ym. olisi teknisesti mahdollista avata kaikkien saataville jo nyt. Arkeologisen kulttuuriperinnön ontologia mahdollistaisi valtakunnallisen muinaisjäännösrekisterin kehittämisen avoimen linkitetyn datan periaatteiden mukaisesti avoimeksi tietoaaineistoksi. Semanttisten sovellusten avulla voitaisiin toteuttaa tehokkaampi tiedonhaku dokumenttien sisällöistä tai esimerkiksi niiden linkittäminen muinaisjäännösrekisteriin. Kenttätutkimuksissa tuotetun sähköisen mittausdatan tallentamista ja hyödyntämistä voitaisiin parantaa merkittävästi; nykyisellään se on käytännössä jälkitöiden valmistuttua hukkaan heitettävää tietoa. Tulevaisuudessa toivottavasti myös aktiivisesti haetaan ja saadaan voimavaroja vanhempienkin tutkimuskertomusten, karttojen, kuvien, päälueellon, verifikaattien ja kortistojen digitointiin.

Arkeologisten kenttätöiden laatuvaatimukset (SALAVA)-hankkeessa on jo testattu käytännössä uudenlaista, yhteisöllistä ja avointa toimintatapaa. Kyseessä ei ole yksittäisen ihmisen tai instituution projekti, vaan arkeologikunnan yhteinen ponnistus. Samalla tavoin on tarkoitus toimia jatkossakin, kun ryhdytään muodostamaan laatuvaatimuksia tukevaa eettistä säännöstöä. Avoimuus ja läpinäkyvyys ovat hyveitä missä tahansa tutkimusprosessissa, ja ne sisällytettäneen myös arkeologian eettisiin periaatteisiin. Luonteva yhteishankkeena toteutettava tema voisi olla myös esimerkiksi tutkimushisto-

riaa, arkeologisen tietämyksemme rajoja ja keskeisiä tutkimuskysymyksiä kartoittava ja esittelevä valtakunnallinen tutkimusohjelma (vrt. Hollannin NOaA; Bazelmans 2006). Tulevaisuudessa henkilökohtaisten tieteellisten meriittien merkitys ehkä vähenee, ja syntyy dynaamisia tutkijakollektiiveja, jotka voivat vaihtaa jäseniä ja tietoja keskenään. Tiedeyhteisöstä tulee aidosti yhteistyössä toimiva, elävä ja muuttuva orgaani. Tieteen avoimuus pakottaa miettimään myös viestinnän tyyliä, tapoja ja välineitä. Sosiaalisen median työkalut helpottavat yhteistyötä, keskustelua ja tietojen jakamista. Verkossa julkaiseminen säästää kustannuksia ja kasvattaa samalla saavutettavuutta merkittävästi paperijulkaisuun verrattuna. Raja populaarisen ja tieteellisen tekstin välillä saattaa hämärtyä. Jokaisen tieteen tekijän tulisikin osata kirjoittaa niin, että kaikki asiasta kiinnostuneet ymmärtävät sisällön.

Avoin tiede tuo tullessaan myös monia ratkaisua vaativia ongelmia: rahoitus, yhteisöprojektien hallinnointi ja koordinointi, tekijänoikeudet, soveltuvat työvälineet – ylipäätään uudenlaisen tieteen mallin sovittaminen vanhoihin rakenteisiin ja käytäntöihin (mm. Harju ja Ropponen 2010). Vaikuttaa kuitenkin siltä, että ongelmat koskettavat ennemminkin yksittäisiä tutkijoita ja tutkimusryhmiä, hyödyt puolestaan koko tiedeyhteisöä ja yhteiskuntaa (Heikkinen 2007). Omaan universumiinsa eristäytyneiden, tietojaan säännöstelevien ja itseriittoisten tieteilijöiden aika on joka tapauksessa ohi. Tieteenharjoittaminen on sosiaalista toimintaa. Avoimuus kysyy kykyä ja halua asettaa itsensä alttiiksi, valmiutta tieteelliseen keskusteluun ja omien näkemysien puolustamiseen. Se on myös suunnaton mahdollisuus oppia muilta, saada tukea ja apua, olla yhdessä enemmän kuin mihin kukaan yksin kykenee.

Kirjallisuus

Arjen tietoyhteiskunnan neuvottelukunta 2010: Tuottava ja uudistuva Suomi – Digitaalinen agenda vuosille 2011–2020. Helsinki. Saatavilla myös verkkojulkaisuna: http://www.arjentietoyhteiskunta.fi/files/320/Digitaalinen_agenda.pdf

Bazelmans, J.G.A. 2006: To What End? To What Purpose? The National Archaeological Research Agenda (NoAa) and Quality Management in Dutch Archaeology. Nationale Onderzoeksagenda Archeologie (version 1.0). BROB 46. Saatavilla verkossa: <http://www.noaa.nl/content/Documenten/1%20Bazelmans%20NOAA%20in%20BROB%2046%20Engels.pdf>

Borg, S. ja Kuula, A. 2007: Julkisasiirtoisten tutkimusdatan avoin saatavuus ja elinkaari. Valmisteluraportti OECD:n datasuositusten toimeenpanomahdollisuuksista Suomessa. Yhteiskuntatieteellisen tietoaikiston julkaisu ja 6. Saatavilla verkossa: http://www.fsd.uta.fi/julkaisut/julkaisusarja/FSDjs06_OECD.pdf

CSC – Tieteen tietotekniikan keskus Oy 2008: Tutkimuksen e-infrastruktuuri Suomessa. Strategiamuistio 17.11.2008. Saatavilla verkossa: <http://www.csc.fi/csc/julkaisut/op-paat>

CSC – Tieteen tietotekniikan keskus Oy 2010: Tutkimuksen tietoaikainistot. Olennaisen käsikirja päättäjille. Helsinki. Saatavilla verkossa: <http://www.csc.fi/csc/julkaisut/op-paat>

Forsström, P.-L. 2011: Tieto käyttöön – Tutkimuksen tietoaikainistot. CSC Tieteen tietotekniikan keskus. Esitys Aalto-yliopistossa järjestetyssä Avoin tieto ja World Wide Web tietoyhteiskunnan palveluksessa -tapahtumassa 5.4.2011. Saatavilla verkossa: <http://www.seco.tkk.fi/events/2011/2011-04-05-yhdistetty-avoin-tietoyhteiskunta/esitykset/Forsstrom.pdf>

Harju, A. ja Ropponen, T. 2010: Tools and Practices for Open Research Collaboration. Short paper presented in The Changing Dynamics of Scientific Collaborations at CSCW, February 7, Savannah GA. Saatavilla verkossa: http://www.sci.utah.edu/images/docs/cscw2010/harju_2.pdf

Heikkinen, M. 2007: Avoin tiede – lyhyt johdatus. Biomi.org. Blogikirjoitus 30.8.2007. <http://www.biomi.org/blogi/2007/08/avoin-tiede-%E2%80%93-lyhyt-johdatus/>

Hyvönen, E. 2011: Miten avoin yhdistetty tieto saadaan verkkoon: kansallinen Linked Data Finland -hanke-ehdotus. Aalto-yliopisto, Semanttisen laskennan tutkimusryhmä. Esitys Aalto-yliopistossa järjestetyssä Avoin tieto ja World Wide Web tietoyhteiskunnan palveluksessa -tapahtumassa 5.4.2011. Saatavilla verkossa: <http://www.seco.tkk.fi/events/2011/2011-04-05-yhdistetty-avoin-tietoyhteiskunta/esitykset/Hyvonen-ldf.pdf>

Hyvönen, E., Viljanen, K., Tuominen, J. ja Seppälä, K. 2008: Building a National Semantic Web Ontology and Ontology Service Infrastructure – The FinnONTO Approach. Teoksessa: Proceedings of the 5th European Semantic Web Conference 2008 ESWC 2008, 110–123. New York. Saatavilla verkossa: <http://www.seco.tkk.fi/publications/2008/hyvonen-et-al-building-2008.pdf>

Kettula, S. 2009: Semanttisen webin ontologisen tekstiilikäsitteistön kehittäminen ja liittäminen museoiden luettelointitietoihin. Kotitalous- ja käsityötieteiden laitoksen julkaisuja 21. Saatavilla verkossa: <http://urn.fi/URN:ISBN:978-952-10-5845-5>

Liikenne- ja viestintäministeriö 2011: Julkisen tietoaikainiston saatavuudesta periaatepäätös. Liikenne- ja viestintäministeriön tiedote 3.3.2011. Saatavilla verkossa: <http://www.valtioneuvosto.fi/ajankohtaista/tiedotteet/tiedote/fi.jsp?oid=322870>

OECD – Organisation for Economic Co-operation and Development 2007: OECD Principles and Guidelines for Access to Research Data from Public Funding. Saatavilla verkossa: <http://www.oecd.org/dataoecd/9/61/38500813.pdf>

Opetus- ja kulttuuriministeriö 2011: Tieto käyttöön. Tiekartta tutkimuksen sähköisten tietoaisteistojen hyödyntämiseksi. Opetus- ja kulttuuriministeriön julkaisuja 2011:4. Saatavilla verkossa: <http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2011/liitteet/okm04.pdf?lang=fi>

Plato 2007: Theaetetus. Translated with an introduction by Benjamin Jowett. Saatavilla verkossa: <http://ebooks.adelaide.edu.au/p/plato/p71th/>

Poikola, A., Kola, P. ja Hintikka, K. A. 2010: Julkinen data – johdatus tietovarantojen avaamiseen. Helsinki. Saatavilla verkossa: <http://www.julkinendata.fi/>

Suonuuti, H. 2006: Sanastotyön opas. Helsinki.

TSK Sanastokeskus 2010: Sosiaalisen median sanasto. Ordlista över sociala medier. Social media vocabulary. TSK 40. Saatavilla verkossa: http://www.tsk.fi/tiedostot/pdf/Sosiaalisen_median_sanasto

Valtioneuvosto 2011: Pääministeri Jyrki Kataisen hallituksen ohjelma 22.6.2011. Saatavilla verkossa: <http://www.valtioneuvosto.fi/hallitus/hallitusohjelma/pdf332889/fi.pdf>

Van Dijk, J. 2006: The Network Society. Social Aspects of New Media (2nd edition). London.

Verkkosivut

Semantic Computing Research Group (SeCo). Aalto-yliopisto. Verkkosivut osoitteessa: <http://www.seco.tkk.fi/>

FinnONTO-hanke. Aalto-yliopisto. Verkkosivut osoitteessa: <http://www.seco.tkk.fi/projects/finnonto/index.fi.php>

Arjen tietoyhteiskunnan neuvottelukunta. Liikenne- ja viestintäministeriö. Verkkosivut osoitteessa: <http://www.arjentietoyhteiskunta.fi/>

CIDOC CRM – International Committee for Documentation (CIDOC) Conceptual Reference Model (CRM). International Council of Museums (ICOM). Verkkosivut osoitteessa: <http://www.cidoc-crm.org/>

CSC – Tieteen tietotekniikan keskus. Opetus- ja kulttuuriministeriö, Koulutus- ja tiedepolitiikan osaston korkeakoulu- ja tiedeyksikkö. Verkkosivut osoitteessa: http://www.csc.fi/index_html

EUDAT – European Data Infrastructure. Verkkosivut osoitteessa: <http://www.eudat.eu/>

Europeana. Verkkosivut osoitteessa: <http://www.europeana.eu/portal/>

JHS-suositukset. JUHTA – julkisen hallinnon tietohallinnon neuvottelukunta. Verkkosivut osoitteessa: <http://www.jhs-suositukset.fi/web/guest>

Kansallinen digitaalinen kirjasto (KDK). Verkkosivut osoitteessa: <http://www.kdk.fi/>

Kulttuurisampo – suomalainen kulttuuri semanttisessa web 2.0:ssa. Verkkosivut osoitteessa: <http://www.kulttuurisampo.fi/>

Linked Data – Connect Distributed Data across the web. Linked Data community. Verkkosivut osoitteessa: <http://linkeddata.org/>

MuseoSuomi – Suomen museot semanttisessa webissä. Verkkosivut osoitteessa: <http://www.museosuomi.fi/>

NOaA – Nationale Onderzoeksagenda Archeologie. Verkkosivut osoitteessa: <http://www.noaa.nl/>

ONKI-ontologiakirjasto. Aalto-yliopisto, Semanttisen laskennan tutkimusryhmä (SeCo) & Helsingin yliopisto. Verkkosivut osoitteessa: <http://onki.fi/fi/>

Protégé-ontologiaeditori. Stanfordin yliopisto. Verkkosivut osoitteessa: <http://protege.stanford.edu/>

TIEKE – Tietoyhteiskunnan kehittämiskeskus ry. Verkkosivut osoitteessa: <http://www.tieke.fi/>

WikiLeaks. Verkkosivut osoitteessa: <http://wikileaks.org/>

Wikipedia – Vapaa tietosanakirja. Verkkosivut osoitteessa: <http://fi.wikipedia.org/wiki/Wikipedia:Etusivu>

World Wide Web Consortium (W3C). Verkkosivut osoitteessa: <http://www.w3.org/>

*FM Johanna Enqvist on työskennellyt Museo-
virastossa tutkijana asiantuntija- ja virano-
maistehtävissä sekä opiskellut viestintää Hel-
singin yliopistossa. Hän suunnittelee parhail-
laan arkeologisen kulttuuriperinnön määrit-
telyä ja merkityksiä käsittelevää väitöskirjaa,
jonka osana on tarkoitus muodostaa myös
arkeologisen kulttuuriperinnön ontologia.
Tutkimusprosessin etenemistä voi seurata ja
kommentoida avoimen tieteen hengessä kir-
joitus- ja ajatustyön tueksi perustettavassa
blogissa.*