

huomiota, jotta voimme luottaa toisiimme paremmin, eikä näistä kysymyksistä sopiminen ole turhanpäiväistä keskustelua vaan kaikki suhtautuvat niihin vakavasti.

Kutsu väärään osoitteeseen

Kutsu kiinteistötoimituksen kokoukseen on mennyt vanhalle omistajalle hänen vanhaan osoitteeseensa. Näinhän ei pitäisi tapahtua, koska Maanmittauslaitoksessa on tiedot kiinteistöjen omistajista ja heidän osoitteistaan. Jos tietoja ajantasaistetaan koko ajan, miten on mahdollista, että edelleen on vanhaa tietoa tallella? Syitä on monia – kukaan tai mikään keskitetty lataus ei ole korjannut virhettä, tieto ei ole tullut perille, tietoa ei ole ehditty korjamaan alkuperäiseen rekisteriin yms. Nämä virheet vähentyisivät, jos aina haettaisiin tieto alkuperäisestä rekisteristä, mutta mikään järjestelmä ei korvaa ihmistä, jonka täytyy tehdä valinta, kun tulee erilaista tietoa eri lähteistä.

Edellä mainitun keskitetyn latauksen osalta on erittäin hankalaa vielä nykyisinkin, mutta valitettavasti vielä ei ole ehditty kehittämään nopeampaa tapaa siirtää tietoja. Toisaalta onko kannattavaa, että valtion ylläpitämässä eri järjestelmissä on samoja tietoja kopioituna. Tuokaan ei kuulosta kovin nykyaikaiselta. Kuitenkin jokainen laitos ja virasto kehittää omia järjestelmiään omista lähtökohdistaan eikä vielä ole kohdattu sitä tilannetta, että kaikilla asiaan liittyvillä organisaatioilla olisi ollut samaan aikaan mahdollisuuksia, rahaa ja voimavaroja kehittää järjestelmiä yhteisesti.

Lopultakin meillä Suomessa ovat asiat suhteellisen hyvin – rekisterit ovat pääsääntöisesti luotettavia ja tiedot niissä oikein.

Kirjoittaja on maanmittausinsinööri, joka työskentelee Maanmittauslaitoksen Kehittämiskeskuksessa. Sähköposti anu.laaksonen@maanmittauslaitos.fi.

”Sähköisessä muodossa säilytettävän tiedon elinkaareen ei yleensä ole kiinnitetty riittävästi huomiota.”

Kykenemmekö kehittämään sähköiseen pitkäaikaissäilytykseen yhtä luotettavia menetelmiä kuin aikanaan kehitetyt savitaulut, papyrukset tai vaikkapa hyvälaatuinen lumpppupaperi?

Marja Rantala

TIEDON SÄILYMINEN luotettavana, eheänä ja käyttökelpoisena on kaiken yhteiskunnallisen ja tieteellisen toiminnan perusedellytyksiä. Tänä päivänä elämme tiedon säilyvyyden osalta epävarmuuden aikana. Teknologian kehittymisen myötä nykyihmisen kyky hallita suuria tietomääriä on yliverlainen verrattuna aikaisempiin aikoihin. Eri tietojärjestelmissä olevien tietojen yhdistelemisen ja käytettävyyden kehittyminen on ollut huimaa muutamiin viimeisten vuosikymmenten aikana. Tietojen säilyvyyden kannalta tilanne ei ole yhtä hyvä eikä kovin yksiselitteinen ja todellisia sähköisen pitkäaikaissäilytyksen ratkaisuja on kehitelty vasta viime



SÄHKÖISEN

aikoina. Kykenemmekö kehittämään sähköiseen pitkäaikaissäilytykseen yhtä luotettavia menetelmiä kuin aikanaan kehitetyt savitaulut, papyrukset tai vaikkapa hyvälaatuinen lumpppupaperi?

Pitkäaikaissäilytys toteutettu analogisilla menetelmillä

Sähköisen pitkäaikaissäilytyksen kehitys on ollut hidasta osin siitä syystä, että varsin pitkään kansalliset arkistolaitokset olivat sillä kannalla, että pysyvään säilytykseen tarkoitettut aineistot säilytetään ainoastaan paperilla tai mikrofilmitulostina ja sähköinen säilyttäminen ei edes



Kuva: Sari Putkonen

PITKÄAIKAISSÄILYTYKSEN HAASTEET

ole todellinen vaihtoehto. Eli pitkäaikais-säilytys on toteutettu analogisilla menetel- millä, vaikka suurin osa tietoaaineistoista on alkujaan ja aktiivivaiheessaan digi- taalista. Käytännössä kuitenkin monista tietoaaineistoista on lähes mahdotonta ottaa pitkäaikais-säilytykseen analogista tulostetta, tyypillisenä esimerkkinä vaika- kapa paikkatietojärjestelmät.

Sähköisen tietoaaineiston myöhemmäl- le käytölle on useita perusehtoja: aineisto on säilynyt fyysisesti, se on luettavissa, käytettävissä, ymmärrettävää ja eheää. Fyysisen säilyvyyden riskejä ovat viruk- set, ohjelmistovirheet ja säilytysmedian eroosio. Luettavuus saattaa estyä sillä, että

uudet laitteistosukupolvet eivät kykene tietoa tulkitsemaan. Tietojen käytettävyy- dessä niiden kaikissa elinkaaren vaiheissa on kysymys tietoaaineiston tunnistami- sesta, esiin ottamisesta ja käsittelystä. Asiakirjallisella tiedolla on elinkaari, joka ulottuu laatimisesta/vastaanottamisesta eri käsittelyvaiheiden kautta pysyvään säilytykseen tai hävittämiseen. Tiedon säilytysmuoto voi vaihdella elinkaaren vaiheissa ja samassakin vaiheessa tiedolla voi olla useampi kuin yksi säilytysmuoto. Sähköisessä muodossa säilytettävän tie- don elinkaareen ei yleensä ole kiinnitetty riittävästi huomiota.

Ymmärrettävyyden kannalta olennai-

nen asia on tiedon konteksti. Kontekstilla tarkoitetaan sitä laajempaa teknistä ja sosiaalista ympäristöä, jonka kanssa asia- kirja on vuorovaikutuksessa. Eheydellä tarkoitamme sähköisen tietoaaineiston aitoutta ja luotettavuutta. Asiakirjallisten tietoaaineistojen kohdalla autenttisuuden varmistaminen on arkistotieteen kannalta olennainen kysymys. Käyttäjän tulee voi- da varmistua dokumentin muuttumatto- muudesta ja aitoudesta, jotta dokumentti olisi todistusvoimainen. Analogisten asiakirjojen kohdalla tiedot alkuperästä ja kontekstista sisältyvät arkistointijärjes- telmään, asiakirjojen fyysiseen paikkaan ja ryhmittelyyn. Sähköisessä pitkäaikais-

säilytyksessä nämä tiedot on tietoisesti tallennettava ja dokumentoitava.

Kaiken kaikkiaan sähköinen tietojen pitkäaikaissäilytys edellyttää jatkuvaa, aktiivista toimintaa säilyvyyden ja käytettävyyden varmistamiseksi ja myös tähän tarkoitukseen varattuja riittäviä ja osaavia resursseja. Olen usein käyttänyt sellaista vertausta, että hyvälle loppupaperille tehdyn asiakirjan voi huoletta ”unohtaa” asianmukaisiin arkistotiloihin sadoiksi vuosiksi ja se säilyy käyttökelpoisena, mutta sähköisessä maailmassa asiakirjojen käytettävyydestä ja säilyvyydestä on pidettävä jatkuvasti aktiivisesti huolta. Tästä jatkuvasta, aktiivisesta säilyttämistehtävästä seuraa automaattisesti myös tietojen arvottamisen tehtävä. Kaikesta syntyvästä tietoaaineistosta on kyettävä seulomaan olennaisimmat osat pysyvään sähköiseen säilyttämiseen, tietojen säilyminen ei voi perustua sattumaan vaan vaatii huolellista harkintaa. Harkintaa ei voida sähköisten

”Käytetyin sähköisen pitkäaikaissäilyttämisen strategia on migraatio. Se tähtää tietoaaineistojen ohjelmisto- ja laitteistoriippumattomaan säilyttämiseen eli digitaalinen aineisto siirretään jaksottain tietystä ohjelmisto- ja laiteympäristöstä tai sukupolvesta toiseen.”

pullinen päätös vastuu tietojen pysyväästä säilyttämisestä on Arkistolaitoksella.

Säilytysstrategiat

Sähköisten tietoaaineistojen pitkäaikaissäilytyksen strategioita ovat mm. museaalinen strategia, emulointi ja migraatio. Museaalinen strategia on nimensä mukaisesti laitteistojen, ohjelmien ja tallennusmedioiden museointia. Atk-museota, josta löytyisivät kaikki mahdolliset tietotekniikan koneet ja ohjelmistot, on kuitenkin käytännössä mahdotonta ylläpitää. Museointi vaatii aina rinnalleen myös kopioinnin eli datan tallenteelta toiselle siirtämisen tietyin aikaväleihin. Eri tallennevälineiden ikä vaihtelee suuresti vuodesta noin 20 vuoteen.

Emulointistrategian mukaisesti sähköisen tietoaaineiston käyttö turvataan ylläpitämällä alkuperäistä käyttöympäristöä. Emulointitekniikasta on suhteellisen paljon tutkimusta ja emulointiin tarvittavia sovelluksia on rakennettu paljon, mutta käytännön toteutuksia vielä suhteellisen vähän. Emulointitekniikan ongelma on mm. se, että jokaisen uuden kone- ja käyttöjärjestelmäsukupolven myötä pitää rakentaa uudet emulaattorit tai vaihtoehtoisesti kasata useita emulaattoreita päällekkäin. Valmiita ohjelmistoja on tarjolla lähimenneisyyden teknisille sukupolville, mutta todellisen pitkäaikaissäilytyksen eli satojen vuosien perspektiivissä tuskin. Emulointitekniikan on ajateltu ratkaisevan mm. multimediaattyyppisten aineistojen ja elektronisten julkaisujen pitkäaikaissäilytyksen ongelmat.

Kaikkein käytetyin sähköisen pitkäaikaissäilyttämisen strategia on migraatio, jonka myös Suomen arkistolaitos ja Maanmittauslaitos ovat valinneet strategiakseen. Migraatiostrategia tähtää tietoaaineistojen ohjelmisto- ja laitteistoriippu-

mattomaan säilyttämiseen eli migraatiossa digitaalinen aineisto siirretään jaksottain tietystä ohjelmisto- ja laiteympäristöstä tai sukupolvesta toiseen. Silloin kun aineisto on jo valmiiksi mahdollisimman ohjelmistoriippumattomassa muodossa kuten esimerkiksi ASCII-tekstinä kyse on vain säilytettävän aineiston siirtämisestä muuttumattomana tallennusmedialta toiselle. Toisessa migraatiotyyppissä vaihdetaan median lisäksi tai sijasta tiedon formaattia eli asiakirjan tallennustapaa muutetaan



Kuva: Ilkka Pietarinen

Maanmittauslaitoksen uuteen arkistokeskukseen Jyväskylään arkistoidaan kaikki uudetkin pysyvään säilytykseen määrättyt maanmittaustoimituksissa syntyneet asiakirjat ja kartat paperisina ja sähköiseen ARKKI-järjestelmään.

tietoaaineistojen osalta useinkaan tehdä enää jälkikäteen vaan se on ratkaistava jo tiedon syntyvaiheessa. Esimerkiksi viranomaisten tietoaaineistoista noin 10–15 % on pysyvä säilytysarvo. Yhteiskunnan perustoimintojen jatkuvuuden kannalta olennaisimpien säilytettävien tietoaaineistojen valinta on suhteellisen helppo tehtävä, mutta tutkimusaineistojen ja -tietokantojen osalta kysymykset ovat huomattavasti monimutkaisempia. Suomen arkistolainsäädännön mukaan lo-

elinkaaren aikana (konvertointi). Migraatio edellyttää jatkuvaa tietotekniikan kehityksen seuranta ja määräväläin tapahtuvaa aineistojen konvertointia.

Standardit sähköisen pitkäaikaissäilytyksen edellytyksenä

Tapahtuipa pitkäaikaissäilytys emulointi- tai migraatiostrategian mukaisesti ovat standardit jota kuinkin yhtä keskeisessä roolissa. Standardien luomista on ajateltu myös varsinaiseksi ratkaisuksi tiedon pitkäaikaissäilytykseen, mutta itsenäiseksi säilyttämisstrategiaksi ne eivät kuitenkaan sovellu. Standardoinnin tilanne vaihtelee hyvin paljon alalta toiselle ja monet standardit ovat lyhytikäisiä ja epävakaita johtuen nopeasta teknisestä kehityksestä. Standardien ikää ja kattavuutta on myös hyvin vaikea ennustaa.

Pitkäaikaissäilytyksen kannalta standardeilla on useita käyttötapoja. Itse tallennusmuotojen standardoinnin rinnalla arkistoinnin kannalta keskeisintä on metatietojen standardointi. Pitkäaikaissäilytyksessä käytettävät tallennusmuodot ovat jo jossain määrin vakiintuneet. ASCII-muoto ja rakenteisten dokumenttien osalta HTML- ja XML-kielet ovat yleisesti käytettyjä. Kuva-aineistojen osalta vakiintunut ja arkistolaitoksen hyväksymä tallennusmuoto esim. TIFF.

Metatietoja käytetään monenlaisiin tarkoituksiin kuten tiedonhakuun ja oikeuksien hallintaan, mutta metatiedoilla on myös hyvin olennainen rooli pitkäaikaissäilytyksessä sekä migraatio- että emulointivaihtoehtoissa. Metatiedoilla varmistetaan sähköisessä pitkäaikaissäilytyksessä nimenomaan asiakirjallisen tiedon (informaatio-objektin) autenttisuus ja alkuperä, tietoaineiston tekninen ja organisatorinen konteksti sekä migraatiohistoria. Pitkäaikaissäilytyksen kannalta olennaisia metatietoelementtejä ovat mm. identifiointitunnus, säilytyshistoria ja oikeudet. Asiakirjallisen tiedon metatietojen määrittelyssä yleisesti hyväksytyt ja käytettyjä standardeja ovat Dublin Core, Model Requirements for the Management of Electronic Records (MoReq), ISO 23081-1 Metadata for records ja JHS143 Asiakirjojen kuvailun ja hallinnan metatiedot sekä erityisesti sähköistä pitkäaikaissäilytystä varten laaditut Arkistolaitoksen SÄHKE-määritykset.

Maanmittauslaitoksen sähköinen arkisto ARKKI

Maanmittauslaitoksessa otettiin viime vuonna käyttöön sähköinen arkistojärjestelmä ARKKI. ARKKI toimii arkistonhal-



Kuva: Ossi Leinonen

Maanmittauspäivien osallistujat tutustuvat ARK:n suunnittelija Hannu Tawastin opastuksella sähköisen arkiston tuotantoon ja erityisesti A0- ja A1-kokosiin tasoskannerilaitteisiin.

lintajärjestelmänä eli se sisältää perusmetatiedot maanmittausarkistoista kattavasti. Metatiedot konvertoitiin uuteen järjestelmään vanhasta sähköisestä erikoisluettelosta eli ELU-järjestelmästä. Metatietoihin liitetään skannatut asiakirjat ja kartat kuvamuodossa. Varsinainen arkistokuva on aina TIFF-muodossa ja palvelukuvat JPEG- tai PDF-muodossa. ARKKI-järjestelmää rakennettaessa yhtenä suurena haasteena oli yhdistää toimiva, nopea sähköinen käyttöarkisto ja tietopalvelutoiminnot sekä sähköisen pitkäaikaissäilytyksen vaatimukset formaatteihin ja metatietoihin. ARKKI-järjestelmää rakennettaessa tehtiin yhteistyötä Arkistolaitoksen SÄHKE-hankkeen kanssa ja tehdyt ratkaisut perustuvat em. hankkeen määrityksiin ja edellä kuvattuihin standardeihin. ARKKI-järjestelmä on kahdennettu ja itse arkistotietovarasto on rakennettu vikasietoiseen levyjärjestelmään, joka sekkin on kahdennettu. ARKKI-tietovarasto kasvaa tämän hetken arvion mukaan noin 4 TB vuodessa. Tietovarasto kasvaa toisaalta uusien toimitusten aineistoilla ja toisaalta aiemmin syntyneen analogisen aineiston systemaattisella, tarveharkintaan perustuvalla skannauksella.

Paperista Maanmittauslaitos ei kuitenkaan vielä lähitulevaisuudessa ole pysy-

väissäilytyksessä luopumassa eli kaikki uudetkin pysyvään säilytykseen määrätty maanmittaustoimituksissa syntyneet asiakirjat ja kartat arkistoidaan paperisena Jyväskylään laitoksen arkistoon. ARKKI-järjestelmä on kuitenkin rakennettu niin, että kokemusten kartuttua voidaan myöhemmin mahdollisesti hakea Arkistolaitoksen lupaa pelkkään sähköiseen säilyttämiseen.



Kirjoittaja on Maanmittauslaitoksen Arkiston johtaja. Sähköposti marja.rantala@maanmittauslaitos.fi.