

Säilyykö tieto ja onko säilynytkään tieto käytettävissä – sähköisen pitkäaikaissäilytyksen haasteet

Marja Rantala

Maanmittauslaitoksen arkisto
Oikokatu 11 40520 Jyväskylä
marja.rantala@maanmittauslaitos.fi

1 Tiedon pitkäaikaissäilytys

Tiedon säilyminen luotettavana, eheänä ja käyttökelpoisena on kaiken yhteiskunnallisen ja tieteellisen toiminnan perusedellytyksiä. Keskeisten asiakirjojen säilyttäminen ja niiden käytettävyyden varmistaminen on aina ollut arkistotoimen perustehtävä, ja arkistoissa oleva asiakirjallinen kulttuuriperintö on se pohja, jolle sukupolvet toisensa jälkeen rakentavat kuvan menneisyydestä eli pohja, jolla toimimme. Varhaisimmat arkistointijärjestelmät syntyivät jo yli tuhat vuotta ennen ajanlaskumme alkua ja olivat hyvinkin kehittyneitä. Säilytysmediana toimi erittäin hyvin säilyvä savitaulu.

Tiedon säilyvyyden ja käytettävyyden kysymykset ovat olleet aina erityisesti arkistotoimesta vastaaville ammatillisesti tärkeitä, haasteellisia kysymyksiä. Arkistollisessa mielessä säilyvyyden ajallinen perspektiivi on todella pitkä, satoja tai tuhansia vuosia. Tietojen eheyden, laadun, käytettävyyden, saatavuuden, säilymisen ja luottamuksellisuuden takaaminen tietojen elinkaaren kaikissa vaiheissa on tärkeää yleisemminkin organisaatioille vaikka aikaperspektiivit ovatkin yleensä lyhyempiä kuin arkistotoimessa.

Tämän päivän maailmassa ja myös suomalaisessa yhteiskunnassa erityisesti tiedon säilyvyyden osalta elämme epävarmuuden aikaan. Teknologian kehittymisen myötä nykyihmisen kyky hallita suuria tietomääriä on ylivoimainen verrattuna aikaisempiin aikoihin. Eri tietojärjestelmissä olevien tietojen yhdistelemisen ja käytettävyyden kehittyminen on ollut huimaa muutamien viimeisten vuosikymmenten aikana. Tietojen säilyvyyden kannalta tilanne ei ole yhtä hyvä eikä kovin yksiselitteinen.

Kykenemmekö kehittämään sähköiseen pitkäaikaissäilytykseen yhtä luotettavia menetelmiä kuin aikanaan kehitetyt savitaulut, papyrukset tai vaikkapa hyvälaatuinen lumppupaperi. Sähköisen pitkäaikaissäilytyksen ratkaisuja on kehitelty vasta viime aikoina. Osin siitä syystä, että varsin pitkään kansalliset arkistolaitokset olivat sillä kannalla, että pysyvään säilytykseen tarkoitettut aineistot säilytetään ainoastaan paperilla tai mikrofilmicom-tulosteina ja sähköinen säilyttäminen ei edes ole todellinen vaihtoehto. Eli pitkäaikaissäilytys on toteutettu analogisilla menetelmillä, vaikka suurin osa tietoaineistoista on alkujaan ja aktiivivaiheessaan digitaalista. Käytännössä kuitenkin monista tietoaineistoista on lähes mahdotonta ottaa pitkäaikaissäilytykseen analogista tulostetta, tyypillisenä esimerkkinä vaikkapa paikkatietojärjestelmät.

Sähköisen tietoaineiston myöhemmälle käytölle on monia perusedellytyksiä: aineisto on säilynyt fyysisesti, se on luettavissa, käytettävissä, ymmärrettävää ja eheää. Fyysisen säily-

vyuden riskejä ovat virukset, ohjelmistovirheet ja säilytysmedian eroosio. Luettavuus saattaa estyä sillä, että uudet laitteistosukupolvet eivät kykene tietoa tulkitsemaan. Tietojen käytettävyydessä niiden kaikissa elinkaaren vaiheissa on kysymys tietoaineiston tunnistamisesta, esiin ottamisesta ja käsittelystä. Asiakirjallisella tiedolla on elinkaari, joka ulottuu laatimises- ta/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 tiedon elinkaaren ei yleensä olla kiinnitetty riittävästi huomiota.

Ymmärrettävyyden kannalta olennainen asia on tiedon konteksti. Kontekstilla tarkoitetaan sitä laajempaa teknistä ja sosiaalista ympäristöä, jonka kanssa asiakirja on vuorovaiku- tuksessa. Asiakirjallisen tietoaineiston ymmärrettävyys irrallaan kontekstistaan on miltei mahdottomuus. Eheydellä tarkoitamme sähköisen tietoaineiston aitoutta ja luotettavuutta. Asiakirjallisten tietoaineistojen kohdalla autenttisuuden varmistaminen on arkistotieteen kan- nalta vaikea kysymys. Käyttäjän tulisi voida varmistua dokumentin muuttumattomuudesta ja aitoudesta, jotta dokumentti olisi todistusvoimainen. Autenttisuus ja kontekstuaalisuus ovat sähköisessä maailmassa tärkeitä arkistotieteen keskeisimmän periaatteen eli provenienssin todentamisessa. Provenienssi viittaa dokumentin taustaan, omistusketjuun ja prosessiin, jonka tuloksena asiakirja/dokumentti on syntynyt. Analogisten asiakirjojen kohdalla tiedot pro- venienssista ja kontekstista sisältyvät arkistointijärjestelmään, asiakirjojen fyysiseen paikkaan ja ryhmittelyyn. Sähköisessä pitkäaikaissäilytyksessä provenienssiin ja kontekstiin liittyvät tiedot on tietoisesti tallennettava ja dokumentoitava.

Kaiken kaikkiaan sähköinen tietojen pitkäaikaissäilytys edellyttää jatkuvaa, aktiivista toimintaa säilyvyyden ja käytettävyyden varmistamiseksi ja myös tähän tarkoitukseen varat- tuja riittäviä ja osaavia resursseja. Tästä jatkuvasta, aktiivisesta säilyttämistehtävästä seuraa automaattisesti myös tietojen arvottamisen tehtävä. Kaikesta syntyvästä tietoaineistosta on kyettävä seulomaan olennaisimmat osat pysyvään sähköiseen säilyttämiseen, tietojen säily- minen ei voi perustua sattumaan vaan vaatii huolellista harkintaa. Harkintaa ei voida sähköis- ten tietoaineistojen osalta useinkaan tehdä enää jälkikäteen vaan se on ratkaistava jo tiedon syntyvaiheessa. Asiakirjallisten tietoaineistojen seulominen ja arvottaminen on sekini ollut jo pitkään keskeisimpiä arkistotoimen tehtäviä, esimerkiksi viranomaisten tietoaineistoista noin 10-15 % on pysyvä säilytysarvo. Yhteiskunnan perustoimintojen jatkuvuuden kannalta olen- naisimpien säilytettävien tietoaineistojen valinta on suhteellisen helppo tehtävä, mutta tutki- musaineistojen ja -tietokantojen osalta kysymykset ovat huomattavasti monimutkaisempia. Suomen arkistolainsäädännön mukaan lopullinen päätösvastuu tietojen pysyvästä säilyttämi- sestä on Arkistolaitoksella.

2 Säilytysstrategiat

Sähköisten tietoaineistojen 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 tal- lennevälineiden ikä vaihtelee suuresti vuodesta noin 20 vuoteen.

Emulointistrategian mukaisesti sähköisen tietoaineiston käyttö turvataan ylläpitämällä al- kuperäistä käyttöympäristöä. Emulointitekniikasta on suhteellisen paljon tutkimusta ja emu- lointiin tarvittavia sovelluksia on rakennettu paljon, mutta käytännön toteutuksia vielä suhteel- lisen vähän. Emulointitekniikan ongelma on mm. se, että jokaisen uuden kone- ja käyttöjärjes- telmäsukupolven 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. multimediatyyppisten aineistojen ja elektronisten julkaisujen pitkäaikaissäilytyksen ongelmat.

Kaikkein käytetyin sähköisen pitkäaikaissäilyttämisen strategia on migraatio, jonka myös Suomen arkistolaitos on valinnut strategiakseen. Migraatiostrategia tähtää tietoaineistojen ohjelmisto- ja laitteistoriippumattomaan säilyttämiseen eli migraatiossa digitaalinen aineisto siirretään jaksoittain tietyistä 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 linkaaren aikana (konvertointi). Migraatio edellyttää jatkuvaa tietotekniikan kehityksen seuranta ja määrävälein tapahtuvaa aineistojen konvertointia. Migraatio ei ole yleisyydestä huolimatta mitenkään yksiselitteinen strategia, vaan pikemminkin kokoelma tiedonsiirtämiseen liittyviä erilaisia lähestymistapoja, joita voidaan muokata eri käyttötarkoituksiin. Migraation soveltaminen edellyttää runsaasti asiantuntemusta ja resursseja, ja migraatiossa tehdyt virheet saattavat olla kohtalokkaita. Migraatioiden huolellinen dokumentointi on ensiarvoisen tärkeää.

3 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 vakiintuneita ja arkistolaitoksen hyväksymiä tallennusmuotoja ovat TIFF ja JPEG.

Metatietoja käytetään monenlaisiin tarkoituksiin kuten tiedonhakuun ja oikeuksien hallintaan, mutta metatiedoilla on myös hyvin olennainen rooli pitkäaikaissäilytyksessä sekä migraatio- että emulointivaihtoehdoissa. Metatiedoilla varmistetaan sähköisessä pitkäaikaissäilytyksessä nimenomaan asiakirjallisen tiedon (informaatio-objektin) autenttisuus ja proveniensi, tietoaineiston tekninen ja organisatorinen konteksti sekä migraatiohistoria. Pitkäaikaissäilytyksen kannalta olennaisia metatietoelementtejä ovat mm. identifiointitunnus, säilytyshistoria ja oikeudet. Asiakirjallisen tiedon metatietojen määrittelyssä yleisesti hyväksytyjä ja käytettyjä standardeja ovat Dublin Core, Model Requirements for the Management of Electronic Records (MoReq) ja JHS143 Asiakirjojen kuvailun ja hallinnan metatiedot. Arkistolaitoksen v. 2001 käynnistyneen SÄHKE-hankkeen (sähköisten tietoaineistojen käsittelyn ja pitkäaikaissäilytyksen kehittämishanke) vielä kesken olevat metatietomäärittelyt tulevat olemaan asiankäsitelyjärjestelmästä sähköisessä muodossa pitkäaikaissäilytykseen arkistolaitokseen siirrettäviä tietoaineistoja koskevia sitovia normeja.