

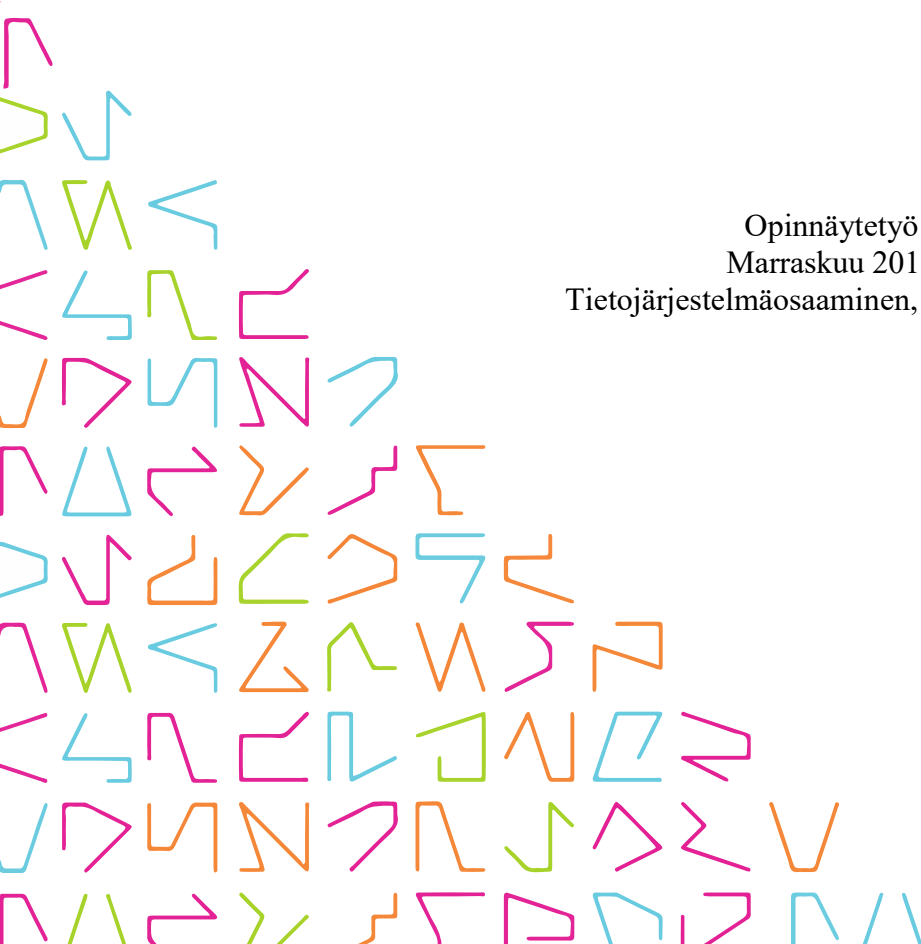


TAMPEREEN
AMMATTIKORKEAKOULU

KOTIMAISTEN KIELTEN KESKUKSEN SÄHKÖISEN AINEISTOKOKONAISUUDEN KEHITTÄMINEN VUOSINA 2015–2019

Toni Suutari

Opinnäytetyö
Marraskuu 2017
Tietojärjestelmäosaaminen, ylempi AMK



TIIVISTELMÄ

Tampereen ammattikorkeakoulu
Tietojärjestelmäosaaminen, ylempi AMK

SUUTARI, TONI:

Kotimaisten kielten keskuksen sähköisen aineistokokonaisuuden kehittäminen vuosina 2015–2019

Opinnäytetyö 79 sivua, joista liitteitä 8 sivua
Marraskuu 2017

Kotimaisten kielten keskus (Kotus) edistää yhteistyössä muiden toimijoiden kanssa kotimaisten kielten dokumentointia ja rakentaa kielitieteellistä tutkimusinfrastruktuuria. Aineistoja tarvitaan kielenkäytön seuraamisessa ja tutkimisessa, ja ne ovat välttämättömiä Kotuksessa tehtävässä kielenhuolto- ja sanakirjatyössä. Aineistot palvelevat myös yhteiskunnan tarpeita.

Sähköinen aineistokokonaisuus tarkoittaa Kotuksen ylläpitämiä kieliaineistojen kokoelmia ja niihin liittyviä palveluja, jotka tarjoavat käyttäjille tutkimus- ja muuta materiaalia esimerkiksi paikannimistä, kirjakielen kehittymisestä tai yleiskielen ja murteiden sanastosta.

Kokoelmia on kartutettu systemaattisesti yli sadan vuoden ajan. Määrät ovat mittavia, ja lisäksi haasteena on aineistojen ja käyttötarpeiden moninaisuus. Lisäksi toimintaympäristö muuttuu kaiken aikaa nopeasti. Toisaalta on tarjolla yhä parempia teknisiä mahdollisuuksia mutta toisaalta käyttäjien odotukset tulevat yhä kovemmiksi. Kehittämistyössä korostuu strategisen ja pitkäjänteyksen suunnittelun merkitys.

Opinnäytteen tarkoituksena on auttaa Kotuksen aineistokokonaisuuden kehittämisessä ja antaa eväitä muun muassa organisaatiotason strategisten valintojen tekemiseen sekä aineistojen kokoelma- ja palvelupolitiikan päivittämiseen. Välittömänä tavoitteena on tarjota näkökulmia aineistoprojektien suunnitteluun ja ohjata olemassa olevan aineistokokonaisuuden uudistumista tietojärjestelmien ja palvelujen osalta. Opinnäyte luo myös kokonais kuvan Kotuksen aineistoninfrastruktuurin historiasta ja nykytilasta sekä strategisista linjauksista. Tapausesimerkkinä kuvataan Nimiarkiston digitalisointihanketta, joka on monessa suhteessa organisaatiolle uudenlainen ja sopii näin malliksi uusille projekteille tai toimintaohjelmille.

Tulevaisuuden skenaarioiden tarkastelussa tunnistetaan riskejä, joita voidaan kohdata, jos toimintaympäristön muutoksiin ei osata tai haluta reagoida ja asettaa riittävän pitkälle ja laajalle ulottuvia strategisia tavoitteita. Innovaation mahdollisuuksia olisi esimerkiksi sanakirja- ja aineistotiedon monipuolisessa yhdistämisessä.

Uudet digitointi- ja muut aineistohankkeet tulee sovittaa määriteltyihin tietojärjestelmä- ja palveluympäristöihin ja välttää aineistokohtaisia erillisratkaisuja.

Asiasanat: aineistot, arkistot, digitalisointi, strategiatyö, tutkimusinfrastruktuuri

ABSTRACT

Tampereen ammattikorkeakoulu
Tampere University of Applied Sciences
Master's Degree Programme in Information Systems Competence

SUUTARI, TONI:

Development of Digital Data Infrastructure in the Institute for the Languages of Finland in 2015–2019

Master's thesis 79 pages, appendices 8 pages
November 2017

The Institute for the Languages of Finland, in cooperation with other parties, promotes the documentation of domestic languages in Finland and maintains linguistic research infrastructure. The data is needed in the research of linguistic usage and it is also essential in the language planning and the dictionary compiling in the Institute. The data infrastructure also serves the needs of the Finnish society.

The term digital data infrastructure refers here the Institute's collections of linguistic data and the services related to them. The services offered to users include material for research and other purposes, for example on place names, development of literary language or vocabulary of standard language and dialects.

The collections have been accumulated systematically over a period of more than a hundred years. The collections are expansive, and the versatility of the data and user needs also poses a challenge. Together with a changing environment, which on one hand offers better technical possibilities, but on the other hand sets higher expectations, emphasizes the significance of strategic and long-term planning.

The intention of this thesis was to help in developing the Institute's data infrastructure and provide ideas for making strategic choices on organizational level and for updating the collection and service policies. The immediate goal was to offer viewpoints for planning data projects and guide the renewal of the existing information systems and services.

The thesis also gives a general view on the history and current state of the Institute's data infrastructure and strategic plans. The digitalization project of the Names Archive is discussed here as a case study that presents a new type of a project for the organization, and is therefore a suitable model for new projects and larger developing programs.

When looking at future scenarios, possible risk scenarios are examined where the changes in working environment are not reacted to and adequate strategic goals are not made for extensive long-term planning. For example, combining dictionary and other data in a versatile way offer possibilities for new innovations.

New digitalization and data projects should be adapted to defined information system and service environments and separate data specific solutions should be avoided.

Key words: data, archives, digitalization, strategic work, research infrastructure

SISÄLLYS

1	JOHDANTO.....	6
1.1	Tarkoitus ja tavoitteet	6
1.2	Menetelmät ja rakenne.....	7
1.3	Kotus organisaationa.....	9
2	Kotuksen aineistoinfrastruktuuri	12
2.1	Aineistot ja niiden ylläpitäminen	12
2.2	Historiaa.....	15
2.3	Nykypäivää	18
2.4	Aineistotyön prosessit.....	20
2.5	Palvelujen kehittäminen.....	22
3	Aineistotyön tekijöiden näkemykset	26
3.1	Kyselyn toteuttaminen	26
3.2	Kyselystä saadut vastaukset ja niiden analysointi	28
3.2.1	Yleiset näkemykset	28
3.2.2	Aineistotyöhön liittyvät näkemykset	30
4	Digitaalisiin kieliaineistoihin liittyvät linjaukset	33
4.1	Kotuksen strategisia linjauksia	33
4.2	Päätös avoimesta lisensoinnista	35
4.3	Visio ja painopisteet.....	36
4.3.1	Visio	36
4.3.2	Aineistojen painopisteet	37
4.3.3	Aineistojen kategoriat	37
4.3.4	Aineistojen käyttöön saattaminen	38
4.4	Kokonaisarkkitehtuuri	38
5	Aineistoihin liittyvää juridiikkaa.....	44
5.1	Lainsäädäntöperusta.....	44
5.2	Tietosuojasta	45
5.3	Tekijänoikeudesta	48
6	Nimiarkiston digitalisointi.....	51
6.1	Nimiarkisto	51
6.2	Alkuvaiheet	52
6.3	Ostopalvelujen hankinta	56
6.4	Osallistaminen	57
6.5	Palvelut asiakkaille	57
6.6	SWOT	58

6.7 Nimilippujen henkilötiedot	60
7 Tulevaisuuden skenaariot	62
8 Päätelmät	65
LÄHTEET	68
LIITTEET	72
Liite 1. Termit ja lyhenteet	72
Liite 2. Kieliaineistojen kokonaisarkkitehtuuri. Luettelo ohjaavasta tiedosta	78

1 JOHDANTO

Kotimaisten kielten keskuksen (Kotus) sähköisen aineistokokonaisuuden kehittäminen kuuluu työtehtäviini, joten oli luontevaa valita aineistojen kehittäminen myös tietojärjestelmäosaamisen ylempään ammattikorkeakoulututkintoon johtaviin opintoihin kuuluvaksi kehittämistehtäväksi. Sähköisellä aineistokokonaisuudella tarkoitan Kotuksen ylläpitämiä kieliaineistojen kokoelmia ja niihin liittyviä palveluja, jotka tarjoavat käyttäjille tutkimus- ja muuta materiaalia esimerkiksi paikannimistä, kirjakielen kehittymisestä tai yleiskielen ja murteiden sanastosta.

Kotuksen aineistot muodostavat kansallisesti merkittävän tutkimusinfrastruktuurin ja ne ovat osa Suomen aineetonta kulttuuriperintöä. Aineistoissa on paljon sellaista sisältöä, joka kiinnostaa tutkijoiden ohella myös muita käyttäjäryhmiä, vaikkapa sukututkijoita ja kotiseutuharrastajia.

Kokoelmia on kartutettu systemaattisesti yli sadan vuoden ajan. Määrät ovat mittavia, ja lisäksi haasteena on aineistojen ja käyttötarpeiden moninaisuus. Lisäksi toimintaympäristö muuttuu kaiken aikaa nopeasti. Toisaalta on tarjolla yhä parempia teknisiä mahdollisuuksia mutta toisaalta käyttäjien odotukset tulevat yhä kovemmiksi. Kehittämistyössä korostuu strategisen ja pitkäjänteisen suunnittelun merkitys.

1.1 Tarkoitus ja tavoitteet

Tämän opinnäytteen tarkoituksena on auttaa Kotuksen aineistokokonaisuuden kehittämisessä ja antaa eväitä muun muassa organisaatiotason strategisten valintojen tekemiseen sekä aineistojen kokoelma- ja palvelupolitiikan päivittämiseen. Välittömänä tavoitteena on tarjota näkökulmia aineistoprojektien suunnitteluun ja ohjata olemassa olevan aineistokokonaisuuden, tietojärjestelmien ja palvelujen uudistumista.

Sähköisiä aineistoja ja niihin liittyviä palveluja Kotuksella on ollut pitkään, joten samanaikaisesti toisaalta kehitetään ja ylläpidetään olemassa olevaa sähköisten aineistojen kokonaisuutta ja toisaalta pyritään siirtämään digitaaliseen muotoon laajoja paperimuotoisia ja muita analogisia arkistokokoelmia.

Tämä opinnäyte toimii toisaalta käytännön työvälineenä kehittämistyössä, kun jatketaan siirtymistä itse tekemisestä yhteisten ratkaisujen käyttämiseen ja digitalisointiin liittyvien palvelujen hankkimiseen kilpailutettuina ostopalveluina.

Vuosikymmenten kuluessa syntyneet sähköiset aineistot ja palvelut eivät muodosta järjestelmien ja palvelujen näkökulmasta ehyttä kokonaisuutta. On aineistokohtaisesti räätälöityjä erillISRatkaisuja, joiden ylläpitäminen on raskasta ja käyttäminen hankalaa. Osa tiedosta on vanhentuneissa järjestelmissä, ja osaa tiedosta ei ole voitu tarjota käyttöön sopivan tietojärjestelmän puuttuessa. Kehittämisen näkökulmasta on olennaista tunnistaa, millaisia tietokokonaisuuksia kieliaineistot muodostavat, ja määritellä, millaisina sähköisinä palvelukokonaisuuksina aineistot halutaan tarjota käyttäjille. Kehittämisessä tulee kiinnittää erityistä huomiota aineistojen yhteiskäyttöisyyteen ja standardinmukaisuuteen.

Organisaation ja sidosryhmien piirissä suhtautuminen keskeisiin muutoksen kohteena oleviin asioihin, kuten tiedon avoimuuteen ja käyttäjien osallistamiseen, vaihtelee jyrkästikin. Eri näkemysten analysoinnin kautta on hahmotettavissa näihin muutoksiin liittyviä riskejä ja mahdollisuuksia ja tätä myötä suunniteltavissa tarkoituksenmukaiselta vaikuttavia toimintamalleja.

Vuosina 2014–2017 on Kotuksessa ollut meneillään yhden ison arkistokokonaisuuden, Nimiarkiston digitalisointiprojekti. Samaan aikaan on kehitetty koko organisaation projektitoimintaa. Olen ollut projektipäällikkönä sekä Nimiarkiston digitalisoinnissa että projektikäsikirjan laatimisessa, joten kumpikin projekti on ollut kiinteä osa kehittämistehtävääni. Nimiarkiston digitalisointi on ollut kokoluokaltaan ja pääosin ostopalveluihin perustuvana Kotukselle uudenlainen toteutustapa, jossa on erityisesti ollut hyötyä projektimaisesta toteutustavasta. Nimiarkiston digitalisointi on tuonut organisaatiolle arvokasta kokemusta sekä vastaavia digitalisointeja että ylipäättään projekteja silmällä pitäen.

1.2 Menetelmät ja rakenne

Opinnäytteen laatimisessa kirjalliset dokumentit ovat olleet keskeisessä osassa. Tämä aineisto voidaan jakaa kolmeen ryhmään. Kotuksen ulkopuolella tuotettujen dokumenttien perusteella olen luonut kuvaa siitä, millainen on toimintaympäristö, mitä mahdollisuuksia

on olemassa ja mitä odotuksia Kotuksen toimintaan kohdistuu. Kotuksen omien dokumenttien avulla olen tarkastellut sitä, miten organisaatio näkee toimintaympäristön ja reagoi siihen. Kolmannen ryhmän muodostavat kehittämistehtävän ja opinnäytteen yhteydessä syntyneet dokumentit, joissa huomio usein kohdistuu ulkoisen toimintaympäristön ja Kotuksen väliseen ”rajapintaan”, siihen miten Kotus kykenee hyödyntämään aineistopalvelujen digitalisoinnissa uusia teknisiä mahdollisuuksia ja tarjoamaan käyttäjien tarvitsemia palveluja.

Käyttämiini menetelmiin kuuluvat myös aineistoyksikön työntekijöille tekemäni Webropol-kysely. Olen myös haastatellut organisaation edellistä johtajaa Pirkko Nuolijärveä sekä aineistoyksikön päällikköä Elisa Stenvallia 9.12.2015. Luonnollisesti kehittämissuunnitelman tekemisessä on tukenut kokemus käytännön työstä, esimerkiksi erilaista palautetta Kotuksen aineistopalveluista tulee usein vaihtelevissa yhteyksissä ja epäsuorasti.

Organisaatiot sitoutuvat moniin asioihin melko lyhyeksi aikaa, joten tulevaisuutta ei voi suunnitella ”yhden kortin varaan”. Lisäksi kehittämisen nopeus ja suunta eivät ole yksinomaan organisaation omissa käsissä, vaan esimerkiksi toimintaympäristöllä ja saatavalla rahoituksella on suuri vaikutus. Tämän vuoksi tulevaisuutta on hyvä tarkastella skenaarioiden kautta ja arvioida, miten organisaation omat tai ulkoiset tekijät voivat joko positiivisesti tai negatiivisesti vaikuttaa siihen, millaiseksi tulevaisuus muodostuu.

Kehittämissuunnitelman pohjaksi tarvitaan sekä nykytilanteen kuvausta että tulevaisuuden visiointia. Pääpaino on kehittämistehtävän aikaisen toiminnan ja tunnistettujen kehittämistarpeiden kuvaamisessa sekä toimintaympäristön odotusten ja vaatimusten tunnistamisessa. Täältä pohjalta muodostan kehittämisen skenaarioita ja arvioin niiden toteutumisen mahdollisuuksia, hyötyjä ja riskejä.

Työn aluksi esittelen Kotuksen aineistoinfrastruktuurin historiaa ja nykytilannetta, samoin tunnistettuja kehittämistarpeita (luku 2). Taustoitusta auttaa ymmärtämään nykytilannetta. Tämän jälkeen esittelen aineistotyössä mukana olevien käsityksiä työstään (luku 3). Digitaalisiin aineistoihin liittyvät linjaukset -luku (luku 4) kuvaa aineistoihin liittyviä strategisia linjauksia, joita olen ollut muotoilemassa osana opinnäytteen taustalla ollutta kehittämistehtävää. Seuraavaksi käyn läpi aineistoihin liittyviä juridisia kysymyksiä (luku 5). Nimiarkiston digitalisointi on esillä omassa luvussaan (luku 6), sillä hanke on

Kotuksen mittakaavassa laaja ja uudenlainen. Näin se toimii hyvänä case-esimerkkinä kehittämisen suunnasta. Lopussa käsittelen tulevaisuuden skenaarioita (luku 7), ja päätelmäluku sisältää kehittämissuunnitelmia (luku 8). Liitteinä mukana ovat termit ja lyhenteet (liite 1) sekä kieliaineistojen kokonaisarkkitehtuurin kuvaukseen kuuluva luettelo ohjauksesta tiedosta (liite 2).

Kehittämistehtävään on liittynyt myös kokonaisarkkitehtuurin näkökulma ja se näkyy tässä opinnäytteessäkin. Osaltaan kokonaisarkkitehtuurin mukaista toimintaa, yhteentömmivuutta sekä yhteisten standardien ja määritysten käyttöä on edistänyt se, että yhä enemmän asioita toteutetaan muistiorganisaatioiden (eli arkistojen, kirjastojen ja museoiden) ja tiedeorganisaatioiden välisenä yhteistyönä eri konsortioissa ja verkostoissa.

1.3 Kotus organisaationa

Kotimaisten kielten keskus eli Kotus on opetus- ja kulttuuriministeriön ohjauksessa toimiva valtion asiantuntijaorganisaatio, josta on säädetty lailla (Laki Kotimaisten kielten keskuksista 1403/2011). Kotus edistää yhteistyössä muiden toimijoiden kanssa kotimaisten kielten dokumentointia ja rakentaa kielitieteellistä tutkimusinfrastruktuuria. Laissa säädettyt tehtävät ovat suomen ja ruotsin kielten huolto, neuvonta ja sanakirjatyö sekä kielenhuoltoon ja sanakirjatyöhön liittyvä tutkimus. Lisäksi tehtävänä on koordinoida saamen kielten, viittomakielten ja romanikielen kielenhuoltoa. Kielitieteen aineistoinfrastruktuurin ylläpitämistä ei laissa mainita, mutta arkistoja ja aineistoja tarvitaan kielenkäytön seuraamisessa ja tutkimisessa, ja ne ovat välttämättömiä Kotuksessa tehtävässä kielenhuolto- ja sanakirjatyössä. Aineistot palvelevat myös yhteiskunnan tarpeita.

Kotimaisten kielten keskuksen strategiassa 2016–2020 (s. 4) kieliaineistoihin liittyvät yleisen tason tavoitteet kuvataan seuraavasti:

”Kotus on keskeinen kieliaineistojen asiantuntija, joka yhteistyössä muiden toimijoiden kanssa edistää kotimaisten kielten dokumentointia ja rakentaa kansallisesti merkittävää kielitieteellistä tutkimusinfrastruktuuria. Aineistotyö tukee Kotuksen ydintoimintoja ja liittyy kansallisesti merkittäviin kielitieteen aineistojen kehittämishankkeisiin. Aineistot ovat osa Suomen aineentonta kulttuuriperintöä.”

”Aineistopalvelut painottuvat suomenkielisten kielen ja kulttuurin kannalta tärkeiden arkistojen ja aineistojen avoimeen tarjoamiseen kansalaisille ja tiedeyhteisölle.”

Kotuksen aineistoinfrastruktuurin ylläpitäminen tarkoittaa pääasiassa kielitieteellistä tutkimusta ja muita kieleen liittyviä tarpeita varten kerättyjen aineskokoelmien ylläpitämistä ja kehittämistä. Tämä kieliaineistojen kokonaisuus sisältää sekä perinteisiä arkistokokoelmia että digitaalisia aineistoja. Digitalisoiminen on kokoelmien kehittämisessä keskeistä. Digitaalisuus koskettaa paperimuotoisia arkistoja vähintään metatietojen osalta, koska jo nyt vähintään perusmetatiedot ovat tietokantamuodossa ja tarjotaan asiakaskäyttöön yhteisen kansallisen ratkaisun, Finna-hakupalvelun kautta.

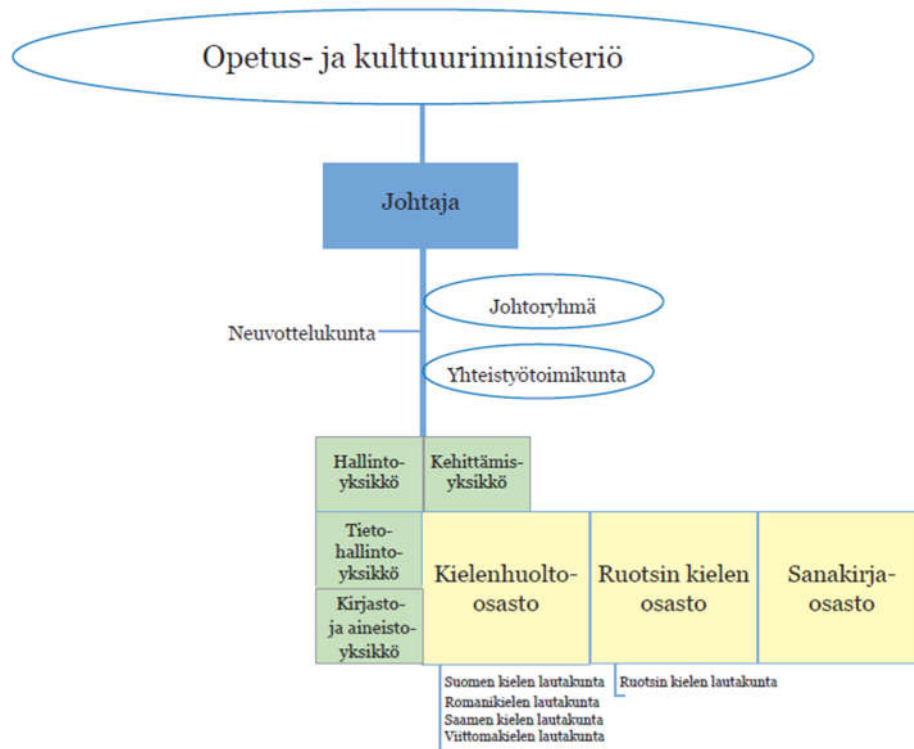
Kotus toimii opetus- ja kulttuuriministeriön hallinnonalalla kieliaineistojen osalta kahdella rintamalla, toisaalta yhdessä muistiorganisaatioiden ja toisaalta yhdessä tiedesektorin kanssa. Kummallakin taholla on viime vuosina voimakkaasti pyritty ministeriön johdolla yhteisiin ratkaisuihin ja alettu laatia tietohallintolain edellyttämiä toimialakohtaisia arkkitehtuurimäärityksiä. Tämän lisäksi on myös muuta toimijoiden välistä yhteistyötä, jossa rakennetaan yhteisiä ratkaisuja ja palveluja.

Kotuksen nykyisessä organisaatiossa substanssitehtävien hoitaminen tapahtuu osastoissa ja tukitoiminnot yksiköissä. Organisaation henkilöstömäärä on noin 80. Kirjasto- ja aineistoyksikössä työskentelee opinnäytteen valmistumishetkellä kuusi henkilöä vakinaisina, heistä neljä aineisto- ja kaksi kirjastotehtävissä. Lisäksi on muutamia määräaikaista projektityöntekijöitä. Aineistotyöhön osallistuvat jossakin määrin myös muiden osastojen ja yksiköiden työntekijät.

Koko organisaation työntekijämäärä on ollut enimmillään yli sadan, ja myös aineistotyön henkilöstöresurssit ovat olleet nykyistä suuremmat. Organisaation rakenne on vaihdellut aikojen saatossa. Seuraavassa kuvassa (kuva 1) näkyy nykyinen organisaatorakenne. Substanssitoiminnasta vastaavat osastot ja tukitoiminnoista yksiköt. Nykyistä kirjasto- ja aineistoyksikköä edelsi aineistoyksikkö. Kirjaston toiminnot siirrettiin kehittämissyksiköstä aineistoyksikköön 15.6.2016.

Kotimaisten kielten keskus

1.8.2016



KUVA 1. Kotimaisten kielten keskuksen organisaatiokaavio. (Kotimaisten kielten keskuksen verkkosivut.)

2 KOTUKSEN AINEISTOINFRASTRUKTUURI

Kotimaisten kielten keskus edistää yhteistyössä muiden toimijoiden kanssa kotimaisten kielten dokumentointia ja rakentaa kielitieteellistä tutkimusinfrastruktuuria. Aineistoja tarvitaan kielenkäytön seuraamisessa ja tutkimisessa, ja ne ovat välttämättömiä Kotuksessa tehtävässä kielenhuolto- ja sanakirjatyössä. Aineistot palvelevat myös yhteiskunnan tarpeita.

2.1 Aineistot ja niiden ylläpitäminen

Kotuksen kokoelmat jakautuvat suhteellisen itsenäisiin arkistokokonaisuuksiin, joista suurimmat ovat Nimiarkisto, Suomen kielen nauhoitearkisto, Suomen murteiden sanaarkisto, Nykysuomen arkisto ja Suomen ruotsalaismurteiden sanaarkisto. Tieteellisistä kokoelmista pääosa on järjestetty tieteellisiä tarkoituksia varten aiheittain (pertinenssi) ja hierarkkisesti. Alkuperään liittyvät tiedot (provenienssi) kuitenkin yleensä sisältyvät metatietoihin tai ovat löydettävissä itse aineistosta. Sen sijaan pienempien, usein yksittäisten tutkijoiden tai hankkeiden tuloksena syntyneiden tieteellisten kokoelmien, (fennistiikan alan) tutkijoiden jäämistöarkistojen ja arkistojen omien hallinnollisten aineistojen järjestämisessä on yleensä käytetty arkistonmuodostajan toimintaa kuvaavaa järjestystä (provenienssiperiaate).

Kokoelmien hallinnan keskeinen työkalu on Kotuksen aineistotietokanta, johon kirjataan aineistojen kuvailevan metadatan lisäksi tekninen ja hallinnollinen metadata ja myös aineiston hallinnan eri vaiheissa tehdyt toimenpiteet. Dokumentoinnista vastaa kunkin aineistokokonaisuuden yhdys- tai vastuuhenkilö. Pyritään mahdollisimman tarkkoihin kuvailutietoihin, mutta käytännössä monet kokoelmat on kuvailtu melko yleispiirteisesti ajatuksena, että kuvailua jatketaan vaiheittain. Kotuksen aineistotietokanta on räätälöity kielitieteellisten tutkimusaineistojen tarpeisiin, mutta sen kehittämisessä on otettu huomioon arkistoalan ja Kansallisen digitaalisen kirjaston käytännöt. Aineistotietokannan viimeaikaisessa kehityksessä on otettu huomioon myös digitaalisen pitkäaikaissäilytyksen metatietovaatimukset.

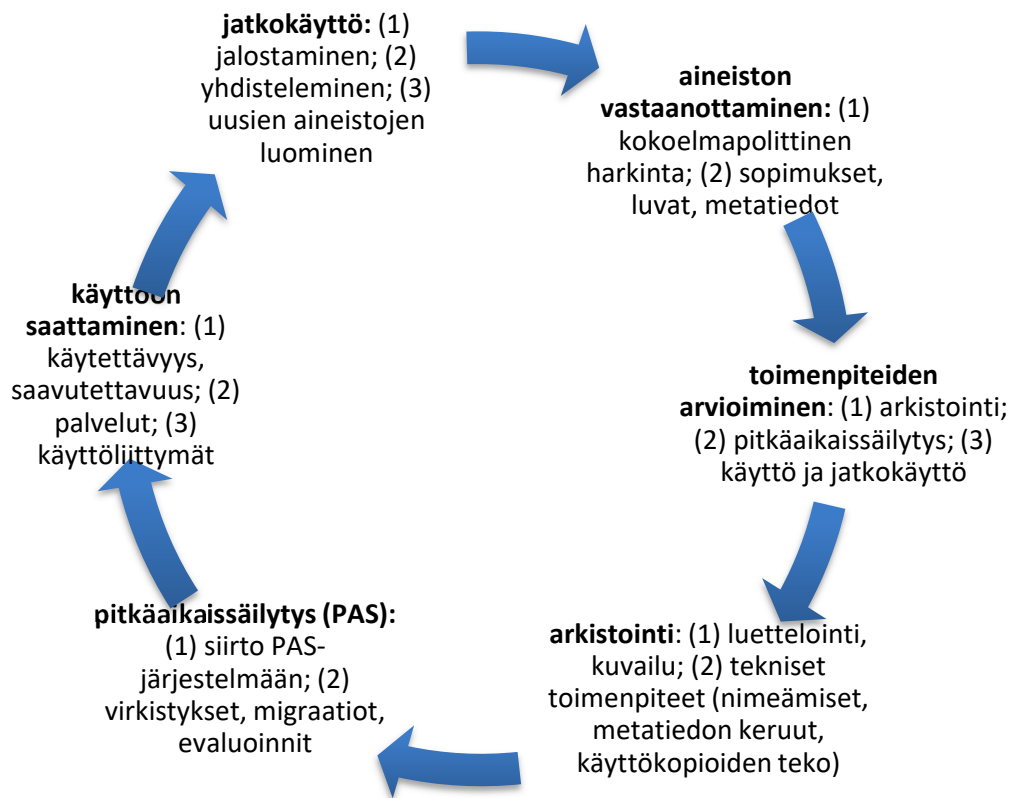
Kotuksen kieliaineistot voidaan jakaa seuraaviin loogisiin tietovarantoihin:

- sana-aineistot (sanaliput, ei-paikkatietoa oleva nimistö, esim. henkilönimet)
- paikkatietoaineistot (paikannimistö, kielen ilmiöiden levinnäisyystiedot)
- audiovisuaaliset aineistot
- tekstikorpukset
- asiakirja-aineistot
- metatiedot

Digitaalisiin kieliaineistoihin liittyy monia rooleja:

- henkilökäyttäjä (tutkija, opiskelija, kansalainen jne.)
- institutionaalinen käyttäjä (hanke, yliopisto, yritys, yhteisö jne.)
- kohdehenkilö (esim. äänitteeseen haastateltu)
- kerääjä/luovuttaja
- tekijänoikeuksien haltija
- palveluntoimittaja (esim. digitointiyritys)
- yhteistyöverkosto/konsortio (KDK/ATT/Fin-Clarín yms.)
- palvelu (Finna, PAS, Kielipankki yms.)

Aineistotyön vaiheita voi hahmottaa elinkaarimallin avulla. Etenkin sähköisten aineistojen ylläpito ja kehittäminen sisältävät paljon aktiivisia toimia edellyttäviä vaiheita, joita alla oleva vuonna 2012 laatimani kuva (kuva 2) havainnollistaa.



KUVA 2. Aineistojen elinkaarimalli Kotuksen aineistojen näkökulmasta (Toni Suutari).

Kotuksen arkistojen ja aineistojen historia alkaa 1800-luvun loppupuolelta, jolloin alettiin systemaattisesti koota murreosanastoa ja nimistöä. Murteiden tallentamiseen äänitteiksi ryhdyttiin laajassa mitassa 1950-luvun lopussa. Vuonna 1976 perustettiin Kotimaisten kielten tutkimuskeskus (vuodesta 2012 nimenä Kotimaisten kielten keskus), jolloin siihen asti eri organisaatioiden ja säätiöiden varassa toimineet arkistot, aineistot, sanakirja- ja tutkimushankkeet sekä kielenhuoltotyö koottiin yhteen.

Yli 100 vuoden aikana koottujen arkistojen laajuus on noin kaksi hyllykilometriä. Sähköisten aineistojen laajuus on satoja miljoonia sanoja, satoja tuhansia kuvatiedostoja ja noin 16 000 tuntia digitaalisia av-tallenteita. Kokoelmia on digitoitu systemaattisesti 1990-luvun lopusta alkaen. Samalla on kehitetty sähköisiä aineistopalveluja.

Sähköisten aineistojen historia juontaa juurensa jo 1970-luvun alkuun, jolloin käytettiin reikäkortteja. Pitkään aineistojen etäkäyttö oli tutkijoille rajattua ja tapahtui VAX- ja UNIX-yhteyksien kautta. (Suutari 2006a: 6–7.)

Nykyaikainen, systemaattinen kokoelmien ja palvelujen digitalisointi on alkanut 1990-luvun lopulla. Vuonna 2007 aineistoja siirrettiin Internetiin selainkäyttöliittymien kautta vapaasti hyödynnettäväksi (Suutari 2006b). Tuolloin toteutus perustui pitkälti omiin ratkaisuihin, vaikka toki esimerkiksi metatiedoissa ja tiedon rakenteessa noudatettiin yleisiä käytäntöjä ja standardeja (esim. DublinCore). Vähitellen digitalisointi on edennyt ja samalla on tullut tarjolle Kielipankin ja Kansallisen digitaalisen kirjaston myötä yhteisiä ratkaisuja. Myös käyttöehtopoliittikka on kehittynyt ja yhä laajemmin aineistot on lisensoitu Creative Commons -lisensseillä. (Suutari 2016.)

2.2 Historiaa

Kotuksen vanhimmat sähköiset aineistot ovat organisaation perustamista edeltävältä ajalta 1970-luvun alusta. Tuolloin Sanakirjasäätiössä lävistettiin reikäkorteille Suomen murteiden sana-arkiston stipendiaattikokoelman tietoja, jotta saatiin tehtyä varsinkin johdosten ja yhdyssanojen tutkimuksessa tarpeellinen käänteisesti aakkostettu sanaluettelo (Tuomi 1975: 23). Tutkimuksen tarpeisiin tehtiin myös muita murteiden ja yleiskielen sanalueltoita. 1970-luvulla päätteenä toimi kirjoitin, joten tulostetut luettelot olivat ainoa tapa hyödyntää aineistoa. Reikäkorteilta aineistot siirrettiin magneettinauhoille ja myöhemmin levyille. Osa on edelleen tallessa ja käytössä, nyt palvelimessa tietokantana.

Kotus aloitti toimintansa 1.3.1976, ja jo seuraavana vuonna perustettiin atk-työryhmä ja palkattiin ensimmäinen tuntipalkkainen atk-suunnittelija. Omia laitteita Kotuksella ei kuitenkaan vielä 1970-luvulla ollut, joten lajittelu- ja tulostustyöt oli kokonaan ulkoistettu Helsingin yliopiston silloiseen laskentakeskukseen. Ensimmäinen oma pientietokone hankittiin vuonna 1980, mutta se kykeni enintään noin 40 000 tietueen suuruisiin lajittelutöihin. (TKS: 11, 13, 15.)

Sähköisiä aineistoja karttui jo 1980-luvun alkupuolella, esimerkiksi tekstikorpusten sanamäärä oli vuonna 1986 pari miljoona. Tekniikan parantuessa kaikki atk-työt keskitettiin Kotukseen ja sähköisten aineistojen käyttömahdollisuudet paranivat. Tietoliikenteen kehitys ja 1988 käyttöön otettu VAX-järjestelmä tehostivat Kotuksen sähköisten aineistojen käsittelyä. Tekstikorpusten etäkäyttö verkon kautta tuli mahdolliseksi 1990-luvun alussa Funetiin liittymisen myötä. Vuonna 1996, kun VAX:n seuraajaksi hankittiin tekstiaineistojen käsittelyyn paremmin sopiva UNIX-järjestelmä. Järjestelmät, joissa palvelimeen

otettiin tekstikomennoin käytettävä pääteyhteys, olivat toimivia ja tarjosivat osaavalle käyttäjälle laajat mahdollisuudet. Moni käyttäjä kuitenkin koki palvelut jo tuolloin hankaliksi. (Suutari 2006a.)

Internet ja selaimella käytettävät palvelut yleistyivät nopeasti ja ohjelmistoissa siirryttiin graafisiin käyttöliittymiin. Internetin läpimurto ajoittui 1990-luvun alkuun, ja varteenotettavaksi suurten aineistojen jakeluvälineeksi se muodostui 1990- ja 2000-luvun vaihteessa. Käyttäjät odottivat helppokäyttöisyyttä ja intuitiivisuutta, myös monen aineiston kohdalla tarpeettomaksi koetusta lupabyrokratiasta luopumista. Kotuksessa muutos tapahtui hitaanlaisesti, sillä vasta vuoden 2007 alkupuolella tekstikorpuksia ja muutamia muita aineistoja avattiin vapaaseen verkkokäyttöön selainkäyttöliittymän kautta. (Suutari 2006b.)

Palvelujen uudistumista edelsi sähköisten aineistojen selvitystyö, joka alkoi kokouksella 9.3.2005. Tuolloin todettiin, että Kotuksella on ”on tarve kehittää aineistojen saavutettavuutta, käytettävyyttä ja dokumentointia”. Selvitettävänä asioina pidettiin muun muassa seuraavia:

- Mitä aineistoja on? Mitä aineistot sisältävät?
- Miten aineistoja pitäisi muokata tai kehittää, jotta ne olisivat julkaistavissa?
- Kuinka kiire aineiston julkaisemisella tai jakeluun saattamisella on?
- Mitkä ovat käyttäjien tarpeet?
- Kuinka paljon ja millaista (sisällöllistä ja teknistä) työtä aineisto vaatii?
- Ketkä ovat tekijät ja vastuuhenkilöt?
- Millainen aikataulu olisi realistinen työn eri vaiheiden toteuttamisessa?
- Voiko aineistoa antaa ulkopuolisten käyttöön (kokonaan tai osittain) esimerkiksi Internetissä tai Kotuksen asiakaskoneissa?
- Miten yksityisyyden suojaan ja tekijänoikeuteen ym. liittyvät seikat vaikuttavat?
- Mikä olisi paras julkaisu- tai jakelutapa (yksittäiset tiedostot, verkkotietokanta Apajassa tai Internetissä jne.)?

Sähköisiä aineistoja käsitelleessä kokouksessa linjattiin alustavasti myös yleisiä periaatteita: ”Suunnitelmissa on muun muassa eräiden tietokantojen saattaminen pikimmiten Kotuksen ulkoisille verkkosivuille selainpohjaiseen käyttöön. Periaatteena on se, että julkaiseen käyttöön annetaan niin paljon aineistoa kuin mahdollista. Niistä aineistoista, joita

ei voida julkaista, voidaan ehkä laittaa näkyviin näytteitä. Verkkoon kootaan aineistoportaali, joka sisältää muun muassa ohjeet ja dokumentaation kaikista Kotuksessa käytettävissä olevista aineistoista.” (Muistio Kotimaisten kielten tutkimuskeskuksen sähköisiä aineistoja käsittelevästä kokouksesta 9.3.2005.) Vuonna 2005 on siis tehty peruslinjaukset, jotka avoimuuden tavoitteen ja dokumentoinnin osalta ovat edelleen ajankohtaiset.

Sähköisten aineistojen selvitystyön tuloksena valmistui maaliskuussa 2006 minun kirjoittamani raportti *Kotimaisten kielten tutkimuskeskuksen sähköiset aineistot*. Siinä jatkettiin kehittämistavoitteiden tunnistamista ja tehtiin ehdotuksia konkreettisista toimista.

- ”Aineistojen määrän kasvaessa on pyrittävä yhä suurempiin kokonaisuuksiin, jotta tiedot ovat järkevällä tavalla löydettävissä ja käytettävissä. Löydettävyyteen liittyy myös aineistorekisterin tarve. Aineistojen käyttöä puolestaan edistävät nykyaikaiset ja helppokäyttöiset www-liittymät sekä turhasta lupabyrokratiasta luopuminen.”
- ”... korostuu pitkäjänteisen suunnittelun tarve, samoin kuin erilaisten standardien ja suosituksen noudattamisen tärkeys.”
- ”Olisi eduksi, jos Kotuksella olisi pitkälle ulottuva *sähköisten palvelujen kehittämissuunnitelma*. Sen perustana olisivat Kotukselle säädetyt tehtävät ja asema niin tiedeyhteisössä kuin yhteiskunnassa laajemminkin. Suunnitelmassa voisi tarkastella sähköisten aineistojen strategiaa laajasti myös vuorovaikutteisesta näkökulmasta, sillä käyttäjien odotukset ja tarpeet kasvavat jatkuvasti samalla kun tekniset mahdollisuudet paranevat. Toisaalta myös muiden tahojen sähköiset palvelut lisääntyvät, joten on syytä miettiä eri toimijoiden välistä työnjakoa.”

Suunnittelutyö jatkui sähköisten aineistojen työryhmässä (4.4.–31.8.2006). Ryhmän lopuraportissa luetellaan keskeisimpinä kehittämiskohteina seuraavat asiat:

- Noudatetaan mahdollisimman avointa lupa- ja käyttöpolitiikkaa. Käyttölupaa vaaditaan vain erityisestä syystä. Pyritään julkaisemaan suurin osa aineistoista Internetissä vapaasti ja helposti käytettävänä.
- Harjoitetaan aktiivista yhteistyötä alan muiden toimijoiden kanssa.
- Nivotaan sähköisiin aineistoihin liittyvät työt aiempaa paremmin osaksi Kotuksen hankkeita ja niiden suunnittelua.
- Perustetaan pysyvä sähköisten aineistojen asiantuntijaryhmä, ja nimetään sähköisten aineistojen vastuhenkilö työskentelemään koordinaattorina.

- Aloitetaan sähköisten sanakirjojen hanke, jonka tavoitteena on julkaista Internetissä *Karjalan kielen sanakirja*, *Vanhan kirjasuomen sanakirjan* ilmestyneet osat ja mahdollisesti myös *Suomen murteiden sanakirja*. Samalla selvitetään myös muiden sanakirjojen (kuten FO, SSA¹) sähköistä julkaisemista.
- Aloitetaan kaikki Kotuksen aineistot ja arkistot kattavan aineistotietokannan kokoaminen.
- Pyritään saamaan hankerahoitusta aineistojen digitalisointiin ja sähköisten aineistopalvelujen kehittämiseen.

(Sähköisten aineistojen työryhmän loppuraportti 31.8.2006.)

Aineistotietokanta on toteutunut ja sen kautta kokoelmien metatiedot on saatu näkymään kirjastojen, arkistojen ja museoiden yhteisessä Finna-hakupalvelussa. Myös arkistojen erillisiä kortistoja on digitalisoitu. Samoin yhteistyö on tiivistynyt ja lupapolitiikka muuttunut avoimeksi. Myös tieteellisten suursanakirjojen ilmestyminen on siirtynyt Internetiin ja painamisesta on luovuttu. Myös tavoite pitkäjänteisestä suunnittelusta, suurista kokonaisuuksista ja kehittämissuunnitelmasta on asetettu jo tuolloin, mutta tältä osin toteutuminen on ollut verkkaisempaa.

2.3 Nykypäivää

Tekstikorpuksat ovat Kotuksen sähköisten aineistojen vanha ydin, joka on syntynyt jo 1980-luvulla. Erityisesti vanhan kirjasuomen ja varhaisnykysuomen korpuksat ovat olleet tutkijoiden aktiivisessa käytössä, ja ovat toki edelleen. Kirjoitetun suomen tekstikorpusten kokoamisessa Kotus ei enää ole aktiivinen, vaan tehtävästä vastaa Kielipankki-konsortio. Tämä on luonteva kehitys, koska uudempien, tekijänoikeuden alaisten tekstien kokoaminen korpuksaksi edellyttää tekijänoikeudellisten kysymysten selvittämistä. Esimerkiksi sopimuslissensijärjestelyistä neuvottelemisen tekijänoikeusjärjestön kanssa sopii hyvin kansallisen konsortion vastuulle.

2000-luvulla Kotuksen sähköisten aineistojen kehittämisessä on ollut keskeistä omien aineistokertaisten aineistojen digitalisoiminen. Systemaattinen analogisen arkistoaineiston di-

¹ FO = *Ordbok över Finlands svenska folkhäl*, Suomen ruotsalaismurteiden sanakirja. SSA = *Suomen sanojen alkuperä*, etymologinen sanakirja.

gitalisointi aloitettiin murreäänitteistä 1990-luvun lopussa. Perusteena oli se, että osa avokelanauhäänitteistä oli tuhoutumisvaarassa. Lisäksi asiakaspalvelusta tuli sujuvampaa, kun kelanauha- tai kasettikopioiden sijasta asiakkaalle voitiin toimittaa äänitteet tiedostoina esimerkiksi cd-levyllä. Kyse oli pitkään nimenomaan tallenteiden digitoinnista. Ratkaisuja digitoitun aineiston verkkokäyttöön on alettu ottaa käyttöön vasta 2010-luvulla.

Äänitteiden digitoinnissa lähdettiin liikkeelle alusta pitäen suunnitelmallisesti ja niin, että katsottiin kokonaisuutta. Muuten monet sähköisiin aineistoihin liittyvät projektit olivat aluksi pieniä ja irrallisina toteutettuja, eli pitkäjänteinen suunnittelu ja kehittäminen eivät toteutuneet vuonna 2006 suunnitellulla tavalla. Olisi ollut eduksi, jos organisaation strategian tasolla olisi riittävän varhaisessa vaiheessa tunnistettu ja valittu aineistojen digitalisoinnin painopistealueet, jolloin voimavarat olisi voitu kohdentaa paremmin. Yleisellä tasolla aineistojen digitalisoimisen tavoite oli selkeästi esillä.

Vuosikymmenten kuluessa syntyneet sähköiset aineistot ja palvelut eivät muodosta järjestelmien ja palvelujen näkökulmasta ehyttä kokonaisuutta. On aineistokohtaisesti räätälöityjä erilliskäyttöjä, joiden ylläpitäminen on raskasta ja käyttäminen hankalaa. Osa tiedosta on vanhentuneissa järjestelmissä, ja osaa tiedosta ei ole voitu tarjota käyttöön sopivan tietojärjestelmän puuttuessa.

Vähitellen on kuitenkin edetty kohti suurempia kokonaisuuksia ja tilannetta ovat parantaneet myös kansalliset hankkeet ja konsortiot, erityisesti Kielipankki ja Kansallinen digitaalinen kirjasto -hanke, joiden myötä on saatu käyttöön yhteisiä alustoja ja järjestelmiä.

Viime vuosina Kotuksessa on siirrytty malliin, jossa hyödynnetään kieliaineistojen monipuolista käyttöä tukevia uusia alan yhteisiä sovelluksia, siirretään verkon kautta käytettäväksi enemmän luvanvaraisia aineistoja, parannetaan metatietojen saavutettavuutta sekä edistetään tietovarantojen avaamista ja hyödynnetään joukkoistamista.

Kieliaineistojen kehittämiseen kokonaisuutena on alettu kiinnittää erityistä huomiota. Vuosina 2013 ja 2014 toteutui Avoin data ja paikkatieto -projekti ja sen jatkoprojekti, joissa paikkatiedon luonteisia aineistoja on kehitetty yhtenäisenä kokonaisuutena ja luotu samalla alustaa paikkatietojen julkaisemiselle (rajapinnat, latauspalvelu) ja käytölle

(haku- ja selailuliittymät, visualisoinnit). Jatkumona tälle aloitettiin vuonna 2014 Nimiarkiston paikannimikokoelmien digitalisointi, jonka kohteena on n. 2,7 miljoonaa suomenkielistä paikannimitietoa ja 10 000 keruukarttaa. Tavoitteena on, että vuoden 2017 lopussa aineisto on keskeisimmiltä osiltaan tietokantamuodossa ja pistekoordinaattipaikannettuna. Hankkeen kokonaiskustannukset ilman oman työn osuutta ovat noin miljoona euroa. Jatkoprojekteiksi on suunniteltu Suomen ruotsin- ja saamenkielisen paikannimistön sisällyttämistä samaan kokonaisuuteen.

Kuvaan ja analysoin muutamia toteutuneita ja menossa olevia aineistojen kehittämiseen liittyviä projekteja. Erityisesti vuosina 2014–2017 toteutuva Nimiarkiston paikannimikokoelmien digitalisointi toimii pilottina, koska siinä toteutetaan asioita uudella tavalla ja koska sen yhteydessä kehitetään uudenlaista palvelukokonaisuutta.

Toimintaympäristö odottaa digitaalisia palveluja, ja digitaalisuus on julkisella sektorilla yksi kehittämisen painopisteistä. Kotuksen näkökulmasta digitaalisuus tarjoaa paitsi keinon lisätä aineistojen käytettyyttä ja vaikuttavuutta, myös uusia mahdollisuuksia osallistaa eri käyttäjäryhmiä niin, että keskeisiä aineistoja voidaan jatkossakin kartuttaa.

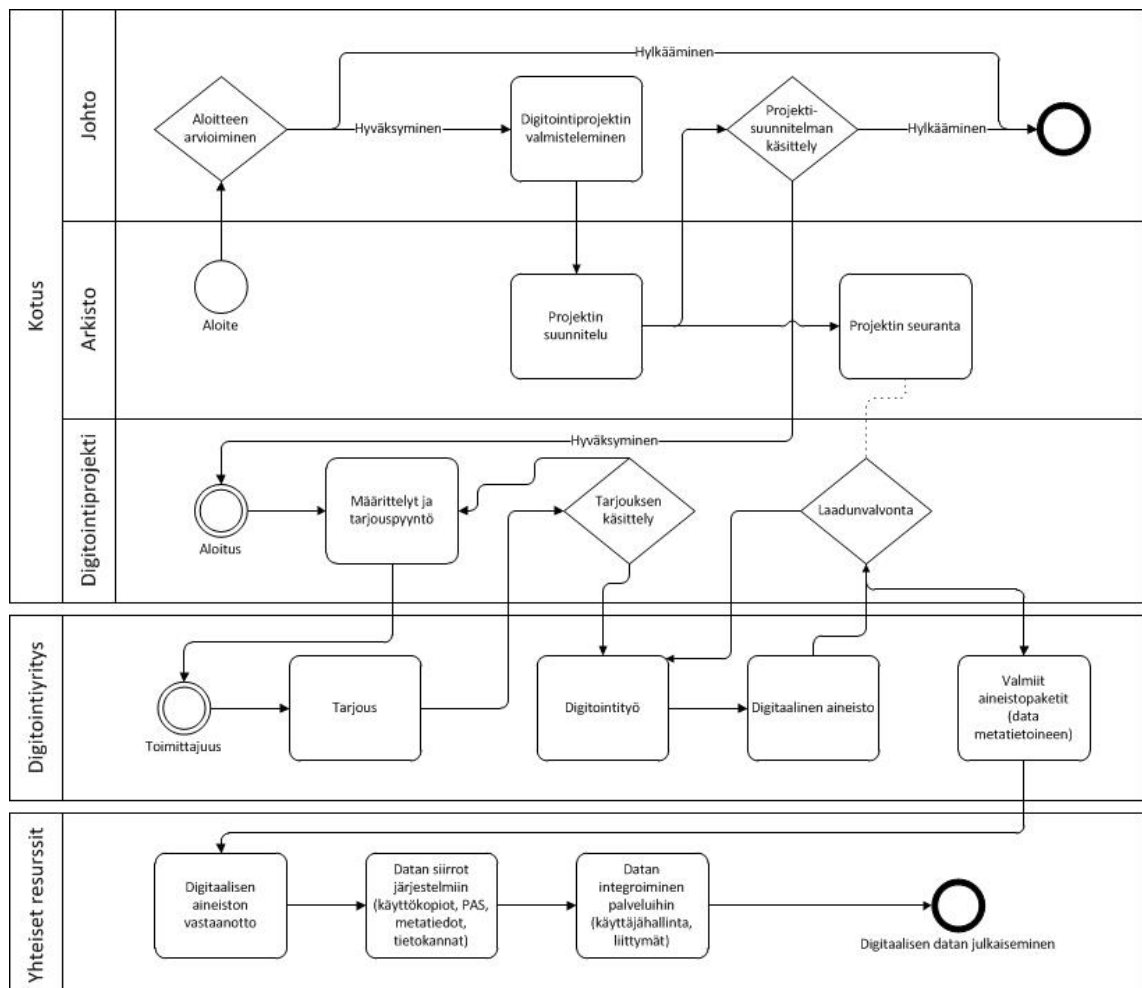
Kotuksen strategiassa kieliaineistoihin liittyvät tavoitteet ovat melko yleisluontoisia (ks. lukua 1.4 *Kotus organisaationa*), mutta kuitenkin pyrkimys muutokseen eli digitaalisiin palveluihin ja vuorovaikutteisuuteen ovat selkeästi esillä.

2.4 Aineistotyön prosessit

Kieliaineistoihin liittyy useita prosesseja, joita on tunnistettu mutta ei juuri kuvattu (ks. kuva 3), sillä prosessikuvauksia Kotuksessa ei ole perinteisesti juuri laadittu. Digitaalisiin kieliaineistoihin liittyvistä prosesseista on kuitenkin kuvattu mm. ostopalveluna hankittavan digitoinnin vaiheet pääpiirteissään (ks. kuva 4).

<p>PALVELUPROSESSIT</p> <ul style="list-style-type: none"> - asiakaspalvelu (kävijä-, sähköposti-, puhelinasiakkaat) - aineiston toimittaminen asiakkaalle kopiona/tiedostona tms. - aineiston digitointi asiakkaalle tilauksesta - käyttöluvan myöntäminen / käyttösovimuksen tekeminen - tietopalvelu - tiedotus <p>KOKOELMIEN HALLINTA</p> <ul style="list-style-type: none"> - aineiston vastaanottaminen ja arkistointi - aineiston käyttöön saattaminen (erit. sähkö. aineistot) - palvelun perustaminen (sähkö. aineistojen käyttöliittymät yms.) - suunnitelmallinen digitointi - pitkäaikaissäilytys - aineiston poistaminen/deponointi <p>HALLINTO</p> <ul style="list-style-type: none"> - ostopalvelun hankkiminen - täydentävän rahoituksen hakeminen - hankkeen elinkaari - työntekijän "elinkaari" (erit. määräaikaaiset) - suunnittelun ja seurannan dokumentit (kertomukset, suunnitelmat, tilastot)

KUVA 3. Kieliaineistoihin liittyviä prosesseja vuodelta 2013. (Toni Suutari.)



KUVA 4. Karkeistettu kuva digitoinnin kokonaisuudesta vuodelta 2015. (Toni Suutari.)

2.5 Palvelujen kehittäminen

Kotus palvelee kieliaineistoillaan varsin heterogeenistä asiakaskuntaa. Monet kokoelmista on kerätty organisaatiossa tehtävän työn tueksi ja muuhun tutkimukselliseen käyttöön, mutta usein kokoelmat palvelevat myös ns. suurta yleisöä, opetusta ja viranomais-toimintaa.

Nykyisin uusien aineistojen karttuminen on resurssien vuoksi hyvin vähäistä ja perustuu suurelta osin tutkijoiden ja tutkimushankkeiden luovutuksiin. Osallistaminen ja talkootyö tarjoavat merkittäviä mahdollisuuksia, joten niitä hyödynnetään muun muassa Nimiarkiston digitalisointihankkeessa ja sen jälkeen nimistötiedon kartuttamisessa ja jalostamisessa. Talkootyössä sinänsä ei Kotuksen näkökulmasta olekaan mitään uutta, sillä esimerkiksi murteiden sanastokokoelmista näet on aikanaan kerätty ”paperiarkistoksi” kansalaisten talkootyönä varsin suuri osa (8 miljoonasta tiedosta noin puolet) ja myös varsin menestyksekkäästi laadukas lopputulos saavuttaen. Nyt on vain löydettävä uudet, digiaikaan sopivat keinot. Osallistamisen myötä vuorovaikutuksessa asiakkaisiin aletaan hyödyntää sosiaalista mediaa ja muita moderneja mahdollisuuksia.

Kotus tekee yhä enemmän kieliaineistoihin liittyvää yhteistyötä eri konsortioissa ja hankkeissa. Yhteiset ratkaisut parantavat eri organisaatioiden tarjoamien aineistojen ja palvelujen yhteentoimivuutta ja luovat kustannustehokkuutta, mutta toisaalta laajat yhteistyökuviot voivat pahimmillaan tarkoittaa kompromisseja ja hidasta edistymistä.

Instituutioiden välisen yhteistyön lisäksi henkilötasolla on monenlaisia yhteyksiä esimerkiksi tutkijoihin, opiskelijoihin ja muihin asiakkaisiin. Asiakasrajapinnassa ja käytännön yhteistyössä mukana oleville organisaation asiantuntijoille kertyy tietoa toimintaympäristön muuttumisesta ja odotuksista. Systemaattisesti kieliaineistoihin liittyvää palautetta ei kuitenkaan pyydetä tai kerätä, vaikka tarve tähän on tiedostettu. Vuodelta 2011 on 28-sivuinen ehdotus Kotuksen palautejärjestelmäksi (*Kotimaisten kielten tutkimuskeskuksen palautejärjestelmä*), ja vuodelta 2012 on 20-sivuinen ehdotus palautejärjestelmän toteuttamiseksi (*Ehdotus Kotuksen palautejärjestelmän toteuttamiseksi*). Toteuttamisehdotuksessa todetaan mm. seuraavaa:

”Palautejärjestelmään kuuluvat palauteprosessit, palautteiden tallentamiseen tarvittava tietojärjestelmä sekä järjestelmän toimijat ja vastuutahot.

Kotuksen yleinen palauteprosessi koostuu palautteen keruusta, käsittelystä ja hyödyntämisestä. Ehdotuksessa kuvataan yksityiskohtaisesti jatkuvan palautteen yleinen prosessi ja lisäksi tarkemmin eräitä kielenhuollon ja sanakirjatyön palauteprosesseja. Palautteet kerätään palauterekisteriin...”

Ehdotuksia palautejärjestelmän toteuttamiseksi ei alettu toteuttaa, joten palautteen keruu ja hyödyntäminen on ollut satunnaista. Nyt asiaan on palattu ja ulkoisen palautteen keräämiseen ryhdytty. Sisäistä palautetta kerätään säännöllisesti mm. työtyytyväisyyskyselyllä ja ”toreilla” sekä henkilökunnalle järjestetyissä keskustelu- ja tiedotustilaisuuksissa. Menneinä vuosina on tehty myös muutama CAF-arviointi².

Työtä ja sen tuloksellisuutta seurataan ja dokumentoidaan suunnittelun ja seurannan asiakirjojen avulla (ks. esim. Kotimaisten kielten keskuksen suunnittelun- ja seurannan asiakirjat). Tuloksellisuutta mitataan myös useiden eri indikaattoreiden avulla. Valtionhallinnon yhteisen Kieku-järjestelmän käyttöönoton myötä valtion henkilöstö- ja taloushallintoa hoidetaan yhtä keskitetympin.

Tietojärjestelmien myötä kieliaineistojen ja niihin liittyvien palvelujen käyttöä ja palveluista saatua palautetta on mahdollista seurata, ja tätä tietoa voidaan käyttää tietojohdatusessa. Edellä kuvatun kaltaisen metatietotason tiedon systemaattiseen keräämiseen ja säilyttämiseen ei kuitenkaan ole mietittyjä prosesseja ja järjestelmiä. Jonkin verran tietoa kerätään vuosittain vuosikertomusten yhteyteen, jolloin on kyse enemmän tietotuotteista.

Kotuksen kieliaineistojen ”metatietotason” tietotuotteina voi pitää tilastoja aineistojen ja käyttäjien määristä sekä sähköisten aineistojen indikaattorilukemaa. Indikaattoreiden avulla myös opetus- ja kulttuuriministeriö seuraa Kotuksen toimintaa.

Kotuksen sähköisten aineistojen indikaattori kuvaa aineistojen määrän ja jalostusasteen kehittymistä kumulatiivisesti. Laskennassa ovat mukana sähköiset aineistot, jotka ovat ulkopuolisen asiakaskunnan käytävissä joko vapaasti tai luvanvaraisesti (ei-kaupallisesti) ja jotka välitetään käyttäjille joko tietoverkoitse tai muuten digitaalisessa muodossa.

² Common Assessment Framework on EU-jäsenmaiden yhteistyönä julkiselle sektorille suunniteltu organisaatioiden laadunarviointimalli.

Kunkin aineiston indikaattoriosuus lasketaan kertomalla sisällöllisten saneiden kokonaismäärän kymmenestuhannesosa aineiston jalostusastetta kuvaavalla kertoimella (1, 2, 10, 11 tai 12). Kerrointen käyttö kuvaa asiantuntijapanosta ja sitä, miten aineistotyö parantaa aineistojen käytettävyyttä, sovellettavuutta ja luotettavuutta. Laskemalla eri aineistojen indikaattoriosuudet yhteen saadaan sähköisten aineistojen kokonaisuutta kuvaava sähköisten aineistojen indikaattorilukema. (Kuva-, ääni ja videoaineistoille lasketaan sane määrä seuraavasti: kartta tai kaavio 1000 sanetta, arkistolippu 20 sanetta, muu kuva 100 sanetta, yksi tunti äänitettä tai videota 6500 sanetta.)

Vaikuttavuutta seurataan jonkin verran esimerkiksi sen perusteella, missä määrin ilmestyy tutkimusjulkaisuja, joiden tekemisessä on hyödynnetty Kotuksen kieliaineistoja. Systemaattinen seuranta ei ole resurssisyistä mahdollista, joten tätä elementtiä ei ole otettu mukaan esimerkiksi sähköisten aineistojen indikaattorin laskentaan. Lisäksi olisi ongelmallista arvottaa erilaisia käyttötapoja, semminkin kun tieteellisen tutkimuksen lisäksi kokoelmilla on opetus-, viranomais- ja kansalaiskäyttöä.

Palvelut ovat aineistojen kannalta keskeisiä. Perinteisissä muodoissa oleviin aineistoihin liittyvät palvelut tarkoittavat hyvin pitkälti perinteistä henkilökohtaista asiakaspalvelua, jossa korostuvat fyysinen ympäristö, kuten tilat ja niiden sijainti, ja henkilökunnan tavoitettavuus ja tietämys.

Digitaalisuus muuttaa palvelut täysin toisenlaisiksi. Näkyvän osan muodostavat käyttöliittymät ja muut käyttöön liittyvät konkreettiset seikat, kuten saatavilla oleva ohjeistus, palvelun tekninen toimivuus ja mahdolliset lupaprosessit. Syvemmällä tasolla joudutaan kuitenkin pohtimaan sitä, muodostuuko sähköisestä tarjonnasta tarkoituksenmukaisia toiminnallisia kokonaisuuksia, joissa on hyödynnetty se lisäarvo, jonka digitaalisuudella ja uusilla viestinnän keinoilla voidaan saavuttaa.

Kotuksen palveluja ei ole yhtenäisesti määritelty, mutta arkiston digitalisoinnin näkökulmasta keskeisiä ovat verkossa tarjottavat aineistojen käyttöön liittyvät palvelut (haku- ja selausliittymät, lataus- ja kyselyrajapinnat). Näiden tueksi tarvitaan opastusta, neuvontaa ja tietopalvelua.

Useinkaan hyvään lopputulokseen ei riitä se, että ”vanhan maailman” ratkaisut siirretään sellaisinaan verkkoon. Käytännössä toki muun muassa saavutettavuus toki paranee, jos vaikkapa arkistolippukokoelmia skannataan ja laitetaan kuvapankkimuodossa verkon kautta selattaviksi, kenties vieläpä jokainen kokoelma vähän eri tavalla teknisesti toteutettuna. Tarvitaan kuitenkin selkeitä ja sisällöltään riittävän laajoja palvelukokonaisuuksia, polkuja, joita asiakas voi edetä haluamallaan tavalla.

Todellista lisäarvoa syntyy usein vasta, kun ymmärretään tarpeet ja mahdollisuudet, esimerkiksi jalostetaan aineistoa niin, että se on haettavissa ja esimerkiksi tietojenkäsittely- ja tilastotieteen menetelmin analysoitavissa. Toisiinsa liittyvien aineistojen linkittäminen ja muu kytkeminen toisiinsa sekä se, että lisensointi ei aseta tarpeettomia käytön rajoituksia, on myös tärkeää. Lisäksi aineiston täytyy olla saatavilla yksittäisen sovelluksen ohella standardinmukaisten rajapintojen kautta. Usein myös käyttäjien osallistaminen ja vuorovaikutteisten ominaisuuksien liittäminen palveluun tuo olennaisia hyötyjä.

Toimintaympäristön muutos on luonut uudenlaisia konsortioita ja yhteishankkeita sekä kansallisella että EU-tasolla, joissa mukana oleminen tuo merkittäviä etuja. Näiden hyötyjen ja vielä hyödyntämättömien mahdollisuuksien tunnistaminen on yksi tavoite, samoin se, missä verkostoissa kannattaa olla mukana ja missä ei.

3 AINEISTOTYÖN TEKIJÖIDEN NÄKEMYKSET

Selvitin Webropol-kyselyllä Kotuksen aineistotyössä mukana olevien käsityksiä työstään. Asiantuntijoiden näkemysten kartoittaminen täydentää kuvaa, joka syntyy organisaation tuottamien dokumenttien pohjalta, ja avaa myös mahdollisuuden vertailla organisaation ja työntekijöiden näkökantoja.

3.1 Kyselyn toteuttaminen

Toteutin Webropolilla anonymin kyselyn 28.10.–3.11.2015. Kysely jakautui yleisesti työhön liittyviä asioita mittaavaan osioon sekä erikseen aineistotyöhön keskittyvään osioon. Pyysin kyselyyn vastauksia Kotuksen aineisto- ja tietohallintoyksikköjen henkilökunnalta. Yleiset kysymykset olivat samoja kuin valtion organisaatioissa käytettävässä työhyvinvointia mittaavassa VMBarossa. Kyselyn teemat olivat seuraavat:

Yleiset kysymykset

- Toimintatavat työyhteisössä (arviointi 1–5)
- Tietoon liittyvät käytännöt (arviointi 1–5)
- Miten hyvin seuraavat väittämät kuvaavat tilannettasi työssä (arviointi 1–5)
- Kuinka monta prosenttia työtehtävistäsi on mieluisia? (%)
- Minkä kouluarvosanan antaisit työpaikallasi kokonaisuutena? arviointi (4–10)

Aineistotyöhön liittyvät kysymykset

- Kotuksen aineistotyön arviointi (arviointi 1–5)
- Omat aineistoihin liittyvät työtehtävät (arviointi 1–5)

Väittämiin vastaaminen tapahtui viisiportaista asteikkoa käyttäen. Aineistotyöhön liittyviin kysymyksiin vastattiin hiukan eri tavalla muotoillulla asteikolla kuin yleisiin kysymyksiin. Pieni luku osoittaa tyytymättömyyttä ja suuri tyytyväisyyttä.

Yleiset kysymykset

- 1 = Ei ollenkaan
- 2 = Hyvin vähän
- 3 = Jonkin verran
- 4 = Paljon
- 5 = Hyvin paljon

Aineistotyöhön liittyvät kysymykset

- 1 = Täysin eri mieltä
- 2 = Eri mieltä
- 3 = Neutraali mielipide / en osaa sanoa
- 4 = Samaa mieltä
- 5 = Täysin samaa mieltä

Asteikkoarviointikysymysten lisäksi pyysin valitsemaan listasta syitä, joiden vuoksi vastaaja pitää työstään, ja asioita, joita vastaaja haluaisi työyhteisössään muuttaa. Valmiiden vaihtoehtojen lisäksi oli mahdollisuus avoimeen vastaukseen. Lisäksi koko kyselyn lopussa oli mahdollisuus vapaaseen kommentointiin.

Aineistoyksikön 6 työntekijästä vastasi 3 ja tietohallintoyksikön 8 työntekijästä (joista 2 Valtorin³ työntekijöitä) vastasi 3. Yksi tietohallintoyksikön työntekijä vastasi aineistosuudessa vain Kotuksen aineistotyön arviointi -osuuteen. Kaikilla vastanneilla on korkeakoulututkinto. Työkokemuksen määriksi ilmoitettiin 8, 20, 22, 24, 25 ja 36 vuotta.

³ Valtion tieto- ja viestintätekniikkakeskus Valtori tuottaa valtionhallinnon toimialariippumattomat ict-palvelut.

3.2 Kyselystä saadut vastaukset ja niiden analysointi

Vastauksen analysoinnissa tarkastelin omina kokonaisuuksinaan yleisiä ja aineistotyöhön liittyviä näkemyksiä. Analysoinnissa myös vertailin kahden eri vastaajaryhmän eli aineistoyksikön ja tietohallintoyksikön työntekijöiden vastauksia.

3.2.1 Yleiset näkemykset

Vastaukset yleisiin kysymyksiin, jotka koskivat toimintatapoja työyhteisössä ja tietoon liittyvät käytäntöjä, tuottivat alle kolmen jäävän keskiarvon. Toimintatavoissa keskiarvoa laskivat erityisesti seuraavat seikat: aloitteiden tekemiseen ei kannusteta, organisaatiossa on heikot etenemismahdollisuudet ja palkitsemisjärjestelmää ei ole. Näissä asioissa kysely ei suoraan mittaa vastaajan työtyytyväisyyttä vaan organisaation luomia edellytyksiä. Esimerkiksi palkitsemisjärjestelmästä lienee hyvin erilaisia mielipiteitä. Joidenkin työhyvinvointia sellaisen käyttö todennäköisesti laskisi ja joidenkin nostaisi. Tietoon liittyvissä käytänteissä niin ikään tiettyjen käytänteiden puute laskee keskiarvoa, esimerkiksi mentoroinnin ja asiakaspalautteen systemaattisen keruun puute sekä se, että uusien ideoiden määrää ei mitata.

Omaan työhön liittyvissä väittämissä vastausten keskiarvo nousee neljään. Valtaosa vastaajista kokee työnsä motivoivaksi, merkitykselliseksi ja arvostetuksi. Lisäksi työtehtävistä koetaan mieluisiksi 70–100 %, ja omalle työhyvinvoinnille annetaan kouluarvosanoja seuraavasti: 7, 8, 8, 9, 9, 9 (n=6). Yksityiskohtaisesti yleisiin kysymyksiin saadut vastaukset on esitetty alla olevassa taulukossa (taulukko 1).

TAULUKKO 1. Kyselyn yleisen osuuden tulokset.

YLEISET KYSYMYKSET		1 Ei ollenkaan	2 Hyvin vähän	3 Jonkin verran	4 Paljon	5 Hyvin paljon	Keski-arvo
Toimintatavat työyhteisössä	Työyhteisössämme tehdään säännöllisesti työtyytyväisyystutkimuksia	0	0	2	2	2	4,00
	Työtyytyväisyystutkimusten tulokset johtavat konkreettisiin toimenpiteisiin	0	2	4	0	0	2,67
	Työyhteisössämme on hyvät mahdollisuudet kehittää ammatillista osaamista	0	0	4	2	0	3,33
	Työyhteisössämme kannustetaan aloitteiden tekemiseen	0	4	0	2	0	2,67
	Työyhteisömmme tarjoaa uralla etenemismahdollisuuksia	0	4	2	0	0	2,33
	Työyhteisössämme käytetään toimivaa palkitsemisjärjestelmää	0	4	2	0	0	2,33
	Johto antaa työntekijöille palautetta säännöllisesti	0	1	5	0	0	2,83
	Työntekijät voivat antaa johdolle palautetta säännöllisesti	0	1	3	2	0	3,17
	Yhteensä	0	16	22	8	2	2,92
Tietoon liittyvät käytännöt	Työyhteisössämme kannustetaan tiedon jakamiseen	0	0	4	2	0	3,33
	Työyhteisössä on käytössä työkiertoa eli henkilöt vaihtavat tehtäviä määräajaksi	2	4	0	0	0	1,67
	Työyhteisössämme käytetään mentorointia	5	1	0	0	0	1,17
	Työyhteisössämme on menossa kehittämisprojekteja/hankkeita	0	1	1	2	2	3,83
	Asiakastytyytyväisyyttä mitataan	1	3	2	0	0	2,17
	Asiakaspalautteet käsitellään ja toimintaa kehitetään niiden pohjalta	0	2	4	0	0	2,67
	Uusien ideoiden määrää mitataan	4	1	1	0	0	1,50
	Kehittämisideoiden esittämiseen on toimivat kanavat	2	1	1	1	1	2,67
	Uutta henkilöstöä perehdytetään systemaattisesti	0	2	1	3	0	3,17
	Poistuvien henkilöiden tiedon jääminen yrityksen käyttöön varmistetaan	0	1	4	1	0	3,00
	Saan tarvitsemaani tietoa tehtävieni (työni) hoitamiseksi	0	0	2	3	1	3,83
	Henkilöstöllä on aikaa töissä oman toiminnan kehittämiseen	0	1	3	2	0	3,17
	Yhteensä	14	17	23	14	4	2,68
	Miten hyvin seuraavat väittämät kuvaavat tilannettasi työssä	Minulla on hyviä työkavereita	0	0	1	3	2
Koen esimieheni arvostavan työtäni		0	0	0	5	1	4,17
Koen esimieheni arvostavan minua		0	0	1	4	1	4,00
Koen kollegoideni arvostavan minua		0	0	1	5	0	3,83
Saan tarvittaessa työhöni apua esimieheltäni		0	0	2	2	1	3,80
Saan tarvittaessa työhöni apua kollegoilteni		0	0	0	4	2	4,33
Työyhteisöni arvot vastaavat omaa arvomaailmaani		0	0	1	4	1	4,00
Koen onnistumisen iloa työssäni		0	0	2	4	0	3,67
Työni motivoi minua		0	0	1	5	0	3,83
Viihdyn työkavereideni seurassa		0	0	0	5	1	4,17
Työni on merkityksellistä		0	0	0	6	0	4,00
Yhteensä		0	0	9	47	9	4,00
SUMMA SUMMARUM	14	33	54	69	15	3,21	

Monissa kohdin aineistoyksikön ja tietohallintoyksikön työntekijöiden vastaukset olivat samansuuntaisia, mutta tietohallintoyksikön keskiarvot jäävät yleensä hiukan alemmiksi kuin aineistoyksikön (taulukko 2). Esimerkiksi aloitteiden tekemisen ja kehittämisideoiden esittämisen mahdollisuuden tietohallintoyksikön vastaajat arvioivat heikommaksi kuin aineistoyksikön vastaajat. Vastaajajoukon pienuus (n=6) voi aiheuttaa vaihtelua sen mukaan, mistä näkökulmasta kukin asiaa katsoo. Tähän voi liittyä se, että yksiköillä on erilainen näkemys siitä, tehdäänkö organisaatiossa säännöllisesti työtyytyväisyystutkimuksia vai ei.

TAULUKKO 2. Kyselyn yleisen osuuden yhteinen keskiarvo, aineistoyksikön vastaajien keskiarvo, tietohallintoyksikön vastaajien keskiarvo sekä yksiköittäin laskettujen keskiarvojen erotus.

YLEISET KYSYMYKSET		Keski-arvo	KA A (aineisto)	KA TH (tietoh.)	KA A&TH erotus
Toimintatavat työyhteisössä	Työyhteisössämme tehdään säännöllisesti työtyytyväisyystutkimuksia	4,00	4,67	3,33	1,34
	Työtyytyväisyystutkimusten tulokset johtavat konkreettisiin toimenpiteisiin	2,67	3,00	2,33	0,67
	Työyhteisössämme on hyvät mahdollisuudet kehittää ammatillista osaamista	3,33	3,33	3,33	0,00
	Työyhteisössämme kannustetaan aloitteiden tekemiseen	2,67	3,33	2,00	1,33
	Työyhteisömme tarjoaa uralla etenemismahdollisuuksia	2,33	2,67	2,00	0,67
	Työyhteisössämme käytetään toimivaa palkitsemisjärjestelmää	2,33	2,33	2,33	0,00
	Johto antaa työntekijöille palautetta säännöllisesti	2,83	3,00	2,67	0,33
	Työntekijät voivat antaa johdolle palautetta säännöllisesti	3,17	2,67	3,67	1,00
	Yhteensä	2,92	3,13	2,71	0,42
Tietoon liittyvät käytännöt	Työyhteisössämme kannustetaan tiedon jakamiseen	3,33	3,33	3,33	0,00
	Työyhteisössä on käytössä työkiertoa eli henkilöt vaihtavat tehtäviä määräajaksi	1,67	1,67	1,67	0,00
	Työyhteisössämme käytetään mentorointia	1,17	1,33	1,00	0,33
	Työyhteisössämme on menossa kehittämisprojekteja/hankkeita	3,83	3,67	4,00	0,33
	Asiakastytyväisyyttä mitataan	2,17	2,00	2,33	0,33
	Asiakaspalautteet käsitellään ja toimintaa kehitetään niiden pohjalta	2,67	2,67	2,67	0,00
	Uusien ideoiden määrää mitataan	1,50	1,67	1,33	0,34
	Kehittämisideoiden esittämiseen on toimivat kanavat	2,67	3,33	2,00	1,33
	Uutta henkilöstöä perehdytetään systemaattisesti	3,17	3,67	2,67	1,00
	Poistuvien henkilöiden tiedon jääminen yrityksen käyttöön varmistetaan	3,00	3,33	2,67	0,66
	Saan tarvitsemaani tietoa tehtävieni (työni) hoitamiseksi	3,83	4,33	3,33	1,00
	Henkilöstöllä on aikaa töissä oman toiminnan kehittämiseen	3,17	3,67	2,67	1,00
	Yhteensä	2,68	2,89	2,47	0,42
Miten hyvin seuraavat väittämät kuvaavat tilannettasi työssä	Minulla on hyviä työkavereita	4,17	4,33	4,00	0,33
	Koen esimieheni arvostavan työtäni	4,17	4,33	4,00	0,33
	Koen esimieheni arvostavan minua	4,00	4,33	3,67	0,66
	Koen kollegoideni arvostavan minua	3,83	4,00	3,67	0,33
	Saan tarvittaessa työhöni apua esimieheltäni	3,80	4,33	3,00	1,33
	Saan tarvittaessa työhöni apua kollegoiltilani	4,33	4,33	4,33	0,00
	Työyhteisöni arvot vastaavat omaa arvomaailmaani	4,00	4,33	3,67	0,66
	Koen onnistumisen iloa työssäni	3,67	4,00	3,33	0,67
	Työni motivoi minua	3,83	4,00	3,67	0,33
	Viihdyn työkavereideni seurassa	4,17	4,33	4,00	0,33
	Työni on merkityksellistä	4,00	4,00	4,00	0,00
	Yhteensä	4,00	4,21	3,76	0,45
	SUMMA SUMMARUM	3,21			

Avoimia vastauksia tuli yleiseen osuuteen vain kaksi. Toinen oli yleinen tyytyväisyyden ilmaus, ja toisessa todettiin, että ”työtyytyväisyyskyselyjen tulosten käsittelyyn pitäisi panostaa enemmän ja muuttaa toiminta palautteen perusteella”.

3.2.2 Aineistotyöhön liittyvät näkemykset

Kotuksen aineistotyön arvioinnissa kriittisimmän arvion saavat resurssien riittävyys ja aineistotyön painoarvo suhteessa organisaation muuhun toimintaan. Silmään pistävää on se, että yksi vastaaja on täysin eri mieltä aineistotyön resurssien oikeasta kohdentamisesta. Selkeästi parhaan arvion saa väittämä siitä, että aineistotyötä arvostetaan asiakkaiden keskuudessa. Myös aineistotyön johtamista asiantuntevasti arvioidaan positiivisesti.

Omiin työtehtäviin vastaajat ovat kyselyn perusteella varsin tyytyväisiä. Heikoimman arvon saa oma halu edetä vaativampiin tehtäviin. Yksityiskohtaiset tulokset on esitetty alla olevassa taulukossa (taulukko 3).

TAULUKKO 3. Kyselyn aineisto-osuuden tulokset.

AINEISTOTYÖHÖN LIITTYVÄT KYSYMYKSET		1 Täysin eri mieltä	2 Eri mieltä	3 Neutr. / EOS	4 Samaa mieltä	5 Täysin samaa m.	Keskiarvo (KA)
Kotuisen aineistotyön arviointi	Aineistotyöllä on selkeät organisaation asettamat strategiset tavoitteet	0	1	2	3	0	3,33
	Aineistotyön painoarvo suhteessa organisaation muuhun toimintaan on oikea	0	3	3	0	0	2,50
	Aineistotyössä on riittävästi osaamista ja asiantuntemusta	0	2	1	3	0	3,17
	Aineistotyöllä on riittävät resurssit	1	3	2	0	0	2,17
	Aineistotyöhön liittyvät osastojen ja yksikköjen välinen työnjako on toimiva	0	1	3	2	0	3,17
	Aineistotyön organisointi on onnistunut	0	0	5	1	0	3,17
	Aineistoyötä johdetaan asiantuntevasti	0	0	1	5	0	3,83
	Aineistotyötä arvostetaan Kotuksessa	0	2	1	3	0	3,17
	Aineistotyötä arvostetaan asiakkaiden keskuudessa	0	0	0	4	2	4,33
	Aineistotyön resurssit on kohdennettu oikein	1	1	4	1	0	2,83
	Yhteensä	2	12	22	22	2	3,17
Omat aineistoihin liittyvät työtehtävät	Minulla on mahdollisuus edetä vaativampiin aineistoihin liittyviin työtehtäviin	0	0	4	1	0	3,20
	Haluun edetä vaativampiin aineistoihin liittyviin työtehtäviin	1	1	3	0	0	2,40
	Saan palautta aineistoyöstäni	0	0	0	5	0	4,00
	Minulla on mahdollisuus esittää aineistoihin liittyviä ideoita ja kehittää toimintaa	0	0	0	4	1	4,20
	Minulla on mahdollisuus kehittää omaa aineistotyöosaamistani	0	0	1	3	1	4,00
	Minulla on riittävästi aikaa tehdä kaikki aineistoihin liittyvät työtehtäväni	0	2	0	3	0	3,20
	Saan tarvittaessa tukea aineistoihin liittyvissä työtehtävissä	0	1	0	4	0	3,60
	Kollegani arvostavat panostani aineistotyössä	0	0	2	3	0	3,60
	Esimieheni arvostaa panostani aineistotyössä	0	0	0	5	0	4,00
	Yhteensä	1	4	10	28	2	3,58
SUMMA SUMMARUM	3	16	32	50	4	3,34	

Suurimmat erot aineisto- ja tietohallintoyksikön vastauksissa liittyvät siihen, kokeeko vastaaja itsellään olevan riittävästi aikaa tehdä kaikki aineistoihin liittyvät työtehtävät (taulukko 4). Aineistoyksikössä tilanne koetaan hyväksi (4,00) kun taas tietohallintoyksikössä huonoksi (2,00).

TAULUKKO 4. Kyselyn aineistotyöhön liittyvien vastausten yhteinen keskiarvo, aineistoyksikön vastaajien keskiarvo, tietohallintoyksikön vastaajien keskiarvo sekä yksiköittäin laskettujen keskiarvojen erotus.

AINEISTOTYÖHÖN LIITTYVÄT KYSYMYKSET		Keskiarvo (KA)	KA A (aineisto)	KA TH (tietoh.)	KA A&TH erotus
Kotuksen aineistotyön arviointi	Aineistotyöllä on selkeät organisaation asettamat strategiset tavoitteet	3,33	3,00	3,67	0,67
	Aineistotyön painoarvo suhteessa organisaation muuhun toimintaan on oikea	2,50	2,33	2,67	0,34
	Aineistotyössä on riittävästi osaamista ja asiantuntemusta	3,17	3,00	3,33	0,33
	Aineistotyöllä on riittävät resurssit	2,17	2,33	2,00	0,33
	Aineistotyöhön liittyvät osastojen ja yksiköiden välinen työnjako on toimiva	3,17	3,33	3,00	0,33
	Aineistotyön organisointi on onnistunut	3,17	3,33	3,00	0,33
	Aineistotyötä johdetaan asiantuntevasti	3,83	3,67	4,00	0,33
	Aineistotyötä arvostetaan Kotuksessa	3,17	3,67	2,67	1,00
	Aineistotyötä arvostetaan asiakkaiden keskuudessa	4,33	4,00	4,67	0,67
	Aineistotyön resurssit on kohdennettu oikein	2,83	2,67	3,00	0,33
	Yhteensä	3,17	3,13	3,20	0,07
Omat aineistoihin liittyvät työtehtävät	Minulla on mahdollisuus edetä vaativampiin aineistoihin liittyviin työtehtäviin	3,20	3,00	3,50	0,50
	Haluan edetä vaativampiin aineistoihin liittyviin työtehtäviin	2,40	2,33	2,50	0,17
	Saan palautta aineistotyöstäni	4,00	4,00	4,00	0,00
	Minulla on mahdollisuus esittää aineistoihin liittyviä ideoita ja kehittää toimintaa	4,20	4,33	4,00	0,33
	Minulla on mahdollisuus kehittää omaa aineistotyöosaamistani	4,00	4,00	4,00	0,00
	Minulla on riittävästi aikaa tehdä kaikki aineistoihin liittyvät työtehtäväni	3,20	4,00	2,00	2,00
	Saan tarvittaessa tukea aineistoihin liittyvissä työtehtävissä	3,60	4,00	3,00	1,00
	Kollegani arvostavat panostani aineistotyössä	3,60	3,67	3,50	0,17
	Esimieheni arvostaa panostani aineistotyössä	4,00	4,00	4,00	0,00
	Yhteensä	3,58	3,70	3,39	0,31
SUMMA SUMMARUM	3,34				

Aineistoihin liittyi myös yksi avoin vastaus, jossa todettiin, että resurssit ovat riittämättömät työn laajuuteen nähden ja että avainhenkilöriippuvuus on liian suurta.

4 DIGITAALISIIN KIELIAINEISTOIHIN LIITTYVÄT LINJAUKSET

Julkisten tietovarantojen avoimuus on globaalisti merkittävä pyrkimys, jota esimerkiksi Euroopan unioni on edistänyt PSI- ja INSPIRE-direktiiveillä⁴ sekä komission avoimen datan strategialla, joka on julkaistu tiedonannossa *Avoim data – innovoinnin, kasvun ja läpinäkyvän hallinnon moottori* vuonna 2011. Kansallisella tasolla julkisten tietovarantojen avaamisen tavoite on sisältynyt Kataisen ja Stubbin hallitusten ohjelmiin, ja digitalisaatio on keskeisessä osassa nykyisin Sipilän hallitusohjelmassa (Kataisen hallitusohjelma 2011, Stubbin hallitusohjelman 2014, Sipilän hallitusohjelma 2015).

Alkuvuodelta 2011 on *Valtioneuvoston periaatepäätös julkisen sektorin digitaalisten tietoa-ineistojen saatavuuden parantamisesta ja uudelleenikäytön edistämisestä*. Toimeenpanon tasolla tietovarantojen avaamisen tavoite on näkynyt muun muassa eri ministeriöiden ohjelmina ja hankkeina, esimerkiksi opetus- ja kulttuuriministeriön hallinnonalalla Tutkimuksen tietoaaineistot (TTA) hankkeena vuosina 2011–2013 ja Avoin tiede ja tutkimus (ATT) -hankkeena vuosina 2013–2017. (Ks. myös Borg & Kuula 2007; Poikola et al. 2010.)

4.1 Kotuksen strategisia linjauksia

Aineistotavoitteita on asetettu Kotuksen toiminta- ja taloussuunnitelmissa sekä opetus- ja kulttuuriministeriön kanssa tehdyissä tulossopimuksissa. Dokumentissa *Opetus- ja kulttuuriministeriön ja Kotimaisten kielten keskuksen välinen tulossopimus vuosille 2017–2020 ja voimavarat vuodelle 2017* todetaan tavoitetilaa käsittelevässä luvussa seuraavaa:

”Kotimaisten kielten keskus (Kotus) lisää kansallista osaamispääomaa asiantuntemuksellaan, avoimella toimintakulttuurilla ja aineistojen laaja-alaisella käyttöön avaamisella mahdollistaen toimintatapojen ja palvelujärjestelmän kehittämisen, uuden liiketoiminnan syntyminen sekä suomalaisen yhteiskunnan osaamiseen perustuvan uudistumisen.”

Lisäksi tulossopimuksessa todetaan erikseen mm., että Kotus vahvistaa yhteiskunnallista vaikuttavuuttaan ”saattamalla aineistojaan kaikkien käyttöön verkossa”. Sopimus sisältää

⁴ PSI-direktiivi koskee julkisen sektorin hallussa olevien tietojen uudelleenikäytön edistämistä ja INSPIRE-direktiivi eurooppalaisen paikkatietoinfrastruktuurin perustamista.

myös viittauksen hallitusohjelman tavoitteeseen ”Digitalisoidaan julkiset palvelut, sujuvoitetaan säädöksiä sekä parannetaan johtamista ja toimeenpanoa”.

Toiminta- ja taloussuunnitelmissa on ollut jo pitkään esillä digitalisaatio. Tietovarantojen avaaminen näkyy erityisesti uusimmissa suunnitelmissa vuosille 2017–2020 ja 2018–2021.

”Tietovarantojen avaaminen: Kotuksen aineistoja pyritään digitoimaan mahdollisimman paljon. Rahoituksen saaminen kuitenkin luo rajat digitoitaville määrille. Digitoidut aineistot lisensoidaan avointa käyttöä varten, ellei tietojen luonne edellytä käyttörajoituksia.”

(Toiminta- ja taloussuunnitelma 2018–2021 s. 9.)

Kotuksen strategiassa korostetaan yhteistyötä muiden organisaatioiden kanssa. Kieliai-
neistoihin liittyvät strategiset valinnat on esitetty seuraavasti:

- ”jatketaan keskeisten aineistojen digitointia ulkopuolisella rahoituksella tai ministeriön erillisrahoituksella ja lisensoidaan ne kattavasti avointa käyttöä varten”
- ”kehitetään ja kartutetaan nimi- ja sana-aineistoja mm. talkoistamista hyödyntäen sekä tarjotaan käyttöön kehittyneinä digitaalisina palveluina mm. osana aineistot ja sanakirjat yhdistävää avoimen tiedon verkkopalvelua”
- ”tarjotaan vähemmän käytettyjä aineistoja digitaalisessa muodossa perusratkaisujen avulla”
- ”pääosa aineistoista tarjotaan CSC:n digitaalisen tiedon säilyttämiseen ja käyttöön liittyvien palvelujen kautta (Kielipankki, AVAA, ETSIN, IDA, PaITuli); kokoelmatiedot ovat asiakaskäytössä Finna-asiakasliittymässä”
- ”huolehditaan kokoelmien luotettavasta säilyttämisestä ja sähköisten aineistojen osalta varaudutaan kansallisen pitkäaikaissäilytysjärjestelmän käyttöönottoon”
- ”pääosa arkistojen asiakaspalvelusta hoidetaan sähköisesti ja paperimuotoiset arkistot ovat käytettävissä erikseen sovittaessa”

(Kotimaisten kielten keskuksen strategia 2016–2020 s. 4.)

Vuodelta 2012 on *Kokoelma- ja palvelupolitiikka: kieliai-
neistot ja kirjasto* -dokumentti (Kokoelma- ja palvelupolitiikka), jossa strategisia tavoitteita on konkretisoitu ja viety

käytännön tasolle kieliaineistojen osalta. Olin mukana politiikkaryhmässä, sen puheenjohtajana. Poliitikassa kiinnitetään strategiaa enemmän huomiota yhteistyöhön, käyttäjien odotuksiin sekä organisaation tarpeisiin hankkia tietoa kehittämistoimintansa tueksi. Poliitiikan päälinjaukset ovat yhteneväiset organisaation strategian kanssa:

- ”Tutkijoiden ja muiden asiakkaiden odotukset kasvavat, kun mm. tilastollisia menetelmiä hyödyntävän tutkimuksen tarpeisiin ja kieliteknologisten sovelusten kehittämisen tueksi tarvitaan pitkälle jalostettuja kieliaineistoja.”
- ”Uudenlaisten aineistopalvelujen kehittäminen onnistuu vain osana kansallisia aineistohankkeita. Alan toimijoiden keskinäinen yhteistyö on entistä tärkeämpää, jotta varmistetaan mahdollisimman laajoja käyttäjäryhmiä palvelevat oikeat digitointiratkaisut ja palvelut.”
- ”Tutkimusmenetelmien ja -kohteiden monipuolistuessa sekä osin muuttuessa-kin ja yhteiskunnan odotusten kasvaessa uusien arkistoaineistojen valikoiva kerääminen vaatii pitkäjänteistä suunnittelua ja tieteenalan ja yhteiskunnan tarpeiden tuntemista.”
- ”Fin-Clarín- ja Kansallinen digitaalinen -konsortiot ovat tärkeitä yhteistyön foorumeita, mutta tiede- ja muistiorganisaatioihin ylläpidetään myös suoria yhteyksiä. Institutionaalisten yhteyksien lisäksi henkilökohtainen verkostoituminen ja aktiivisuutta arvostetaan. Kotuksen aineistotyön tueksi pyritään muodostamaan yliopisto-opettajien ja -tutkijoiden joukko, jonka kautta käytännön yhteistyötä voidaan tehokkaasti edistää, esimerkiksi sisällyttää aineistotyötä osaksi opintoja. [--] Yhteistyötä tehdään myös kansalaisten ja kolmannen sektorin yhteisöjen kanssa, sillä tavoitteena on löytää nykyaikaan sopivia talkoistamisen keinoja...”

(Kokoelma- ja palvelupolitiikka: kieliaineistot ja kirjasto s. 4, 5, 19.)

Kokoelma- ja palvelupolitiikkaakaan ei ole päivitetty vuoden 2012 jälkeen, vaikka alun perin tarkoituksena oli, että dokumentti olisi säännöllisesti päivitettävä. Tarkoituksena on, että kieliaineistojen kokoelma- ja palvelupolitiikka päivitetään vuonna 2018, kun mitava ja resursseja sitova Nimiarkiston digitalisointihanke on saatu päätökseen.

4.2 Päätös avoimesta lisensoinnista

Kotuksen kieliaineistoja koskeviin strategisiin linjauksiin on liittynyt muutamia merkittäviä välietappeja, jotka ovat edistäneet tavoitteiden jalkauttamista. Edellä mainittu

vuonna 2012 laadittu kokoelma- ja palvelupolitiikka sisältää erilaisia käytännön tason ohjeita, kuten periaatteet siitä, missä järjestyksessä digitalisointia edistetään.

Vuoden 2014 lopussa tehtiin esityksestäni päätös, jonka mukaan Kotuksen nykyiset ja tulevat sähköiset kieliaineistot, jotka ovat julkisia ja joiden vapaata käyttöä lainsäädäntö tai sopimukset eivät rajoita, avataan julkisina tietovarantoina vapaasti käytettävänä, Creative Commons -lisenssillä lisensoituina ja koneluettavina. Päätös ei koskenut sanakirja-aineistoja. Samoihin aikoihin julkaistiin JHS 189 Avoimen tietoaineiston käyttöluvan suositus, jonka mukaan Suomen julkishallinnon avoimiin tietoaineistoihin suositeltava käyttöluva on Creative Commons Nimeä 4.0.

4.3 Visio ja painopisteet

Aineistotyön visio ja painopisteet määriteltiin vuonna 2015. Määrittely oli osa kehittämistehtävääni. Pontimena tilanteen arvioimiselle antoi se, että Kotus oli muuttamassa valtion toimitilastrategian ja sisäilmaongelmien vuoksi uusiin tiloihin, ja näytti siltä, että uusiin tiloihin ei voida sijoittaa kaikkia arkistoaineistoja. Ajatuksena oli, että aineistojen laajamittainen digitoiminen vähentää tilantarvetta ja tilakustannuksia. Digitoidut alkupe- räisaineistot olisi voitu sijoittaa etäarkistoon, jolloin olisi syntynyt merkittäviä tilakustannussäästöjä, kun muualla sijaitsevien tilojen vuokrataso on Helsingin keskusta-alueella alempi ja kun aineistot olisi voitu pakata tiiviisti eikä olisi tarvittu asiakastiloja kokoelmien yhteyteen. Laajamittainen digitalisointi ei toteutunut rahoituksen puutteesta, mutta löytyi muita ratkaisuja. Aineistoja deponoitiin muille organisaatioille, ja av-aineistojen säilytys siirrettiin kilpailutetun palveluntarjoajan hoidettavaksi. Lisäksi muutamia toissijaisia aineistoja digitoitiin ja hävitettiin. Ja lopulta uusiin tiloihin saatiin järjestettyä ennakko-odotuksia enemmän säilytyshyllymetrejä.

Joka tapauksessa aineistotyö sai selkeän vision ja painopisteiden määrittelyn, jotka myös organisaation johtoryhmä käsitteli.

4.3.1 Visio

Kotus ylläpitää ja kehittää vain tavoitteidensa kannalta keskeisiä aineistoja. Aineistoista ei haluta vain historiallisia, vaan niitä pitää voida kartuttaa ja kehittää. Digitoinnin myötä

tähän avautuu mahdollisuus, kun voidaan osallistaa tutkijoita, yhteisöjä ja kansalaisia. Voimavarat on kohdennettava selkeästi. Kaikkea mahdollista ei voida ylläpitää.

Muilla kuin keskeisille aineistoille on etsitty parempi sijoituspaikka toisista organisaatioista. Monille Kotuksen toiminnan kannalta vähemmän keskeisille aineistoille on löydettävissä luontevampi sijoituspaikka muista organisaatioista.

Kaikki Kotuksen keskeiset aineistot on digitoitu ja lisensoitu avointa käyttöä varten, ellei tietojen luonne edellytä käyttörajoituksia. Digitointi parantaa saavutettavuutta ja käytettävyyttä sekä säästää tilakustannuksissa. Digitointi luo mahdollisuuksia uudentamalleihin, kuten sanakirjojen ja aineistojen yhdistämiseen sekä toimitustapojen uudistumiseen.

4.3.2 Aineistojen painopisteet

Keskeiset aineistot liittyvät suomeen ja suomenruotsiin, ja nykykielen tallentamista jatketaan kieliperinnön katkeamattoman dokumentoinnin varmistamiseksi.

Aineistot tukevat Kotuksen ydintoimintoja ja ovat kansallisesti merkittävä tutkimusinfrastruktuuri sekä osa Suomen aineetonta kulttuuriperintöä.

4.3.3 Aineistojen kategoriat

Vision ja painopisteiden yhteydessä määriteltiin neljä aineistokategoriaa.

Ydinaineistot: kehitetään ja kartutetaan mm. joukkoistamista hyödyntäen sekä tarjotaan käyttöön kehittyneinä digitaalisina palveluina, jotka tukevat ydintoimintoja eli kielenhuoltoa ja sanakirjatyötä, mm. paikkatiedot (erityisesti paikannimet), sana- ja sanakirja-aineistot, kokoelmien perusmetatiedot ja muut ydintoiminnoille tärkeät aineistot.

Perusaineistot: ei kartuteta, mutta tarjotaan käyttöön digitaalisessa muodossa perusratkaisujen avulla, mm. av-aineistot, tekstikorpukset ja erilliskeräelmät.

Säilytettävät aineistot: ei kartuteta eikä digitoida, paitsi erityisestä syystä tai asiakastilauksen perusteella, mm. etymologiset kokoelmat, toissijaiset erilliskokoelmat ja sekalaiset, muissa kohdissa mainitsemattomat aineistot.

Muualle siirrettävät aineistot: luovutetaan mahdollisuuksien mukaan aineistoista, jotka sopivat paremmin jonkin toisen organisaation kokoelmiin.

4.3.4 Aineistojen käyttöön saattaminen

Digitaalisten aineistojen käyttöön saattamisessa hyödynnetään CSC:n palveluja (AVAA, IDA ym.) ja Kielipankkia (LAT, Korp, Sanat).

Nimiarkiston paikannimikokoelmien digitointihanke toimii pilottina ja sen myötä syntyy ratkaisuja, joita voidaan hyödyntää myös muissa arkistolippumuotoisissa aineistoissa ja paikkatietoaineistoissa.

Perusdigitoinnin jälkeen ydinaineistot vaativat merkittävästi jatkojalostamista ja digitaalisten palvelujen kehittämistä, jotta digitoiminen palvelisi käyttäjien tarpeita ja avaisi uusia hyödyntämismahdollisuuksia. Jatkojalostamisen kustannukset ovat vähintään samaa luokkaa kuin perusdigitoinnin kustannukset.

4.4 Kokonaisarkkitehtuuri

Kokonaisarkkitehtuuri toimii organisaation jatkuvan, iteratiivisen toiminnan kehittämisen välineenä. Lähtökohtina ovat johtaminen ja strategia, jotka ohjaavat sekä kokonaisarkkitehtuurityötä että sen myötä tapahtuvaa kehittämistä. Näin ollen kokonaisarkkitehtuuria ei pidä nähdä pelkästään tietohallintoa vaan koko organisaatiota koskevana asiana, vaikka julkisen sektorin toimijoille juridinen velvoite menetelmän käyttöön onkin määritelty tietohallintolaissa, jonka 7 §:ssä todetaan seuraavaa:

”Julkisen hallinnon viranomaisen on julkisen hallinnon tietojärjestelmien yhteentoimivuuden mahdollistamiseksi ja varmistamiseksi suunniteltava ja kuvattava kokonaisarkkitehtuurinsa sekä noudatettava laadittua ja ylläpidettyä kokonaisarkkitehtuuria ja sen edellyttämiä yhteentoimivuuden kuvauksia ja määrittämiä sekä toimialakohtaisia tietojärjestelmien yhteentoimivuuden kuvauksia ja määrittämiä.”

Julkisella sektorilla kokonaisarkkitehtuuri on keskeinen väline kehitettäessä yhteentoimivuutta koko ”valtiokonsernin” tasolla. Tämä tavoite on keskeinen myös Kotuksen kieliaineistojen kannalta.

Kokonaisarkkitehtuurimallin käyttöönoton edistäminen on ollut osa kehittämistehtävääni. Laadin vuonna 2015 kuvauksen Kotimaisten kielten keskuksen digitaalisten kieliaineistojen kohdearkkitehtuurin nykytilasta (Suutari 2015). Dokumentti sisältää muun muassa arkkitehtuuriperiaatteet, strategiset linjaukset ja strategiset tavoitteet. Kuvaus valmistui ennen kuin edellä esitellyt vision ja painopisteiden määrittelyt oli organisaatiossa käsitelty, joten arkkitehtuurikuvaus tukeutuu silloin voimassa olleeseen strategiaan sekä kokoelma- ja palvelupolitiikkaan.

Arkkitehtuuriperiaatteissa näkyy valtionhallinnon yleinen pyrkimys julkishallinnon tietovarantojen avaamiseen (taulukko 5). Periaatteen lähteiksi on kuitenkin valittu dokumentteja, joissa yleiset valtionhallinnon tavoitteet ja toisaalta lainsäädännön velvoitteet (mm. tietoturva) ja vastaavat asiat on esitetty Kotuksen aineistojen näkökulmasta konkreettisemmalla tasolla.

Digitaalisiin kieliaineistoihin liittyy sekä yleisiä että toimialakohtaisia suosituksia. JHS-suositukset ICT-palvelujen kehittämisestä (JHS 171, 172, 179) ovat keskeisiä, samoin JHS 189 (Avoimen tietoaineiston käyttöluva), joka on suositus avoimen tietoaineiston CC-lisensioimisesta. JHS-suositukset hyväksyy julkisen hallinnon tietohallinnon neuvottelukunta JUHTA.

Kielitieteen tutkimusinfrastruktuurin ylläpitäjänä Kotus on mukana sekä muistiorganisaatioiden että tiedesektorin verkostoissa, joista jokainen osaltaan luo yhteisiä ratkaisuja ja asettaa de facto -reunaehdoja.

Muistiorganisaatioiden eli kirjastojen, arkistojen ja museoiden kannalta keskeinen on opetus- ja kulttuuriministeriön hallinnonalalla toimiva Kansallinen digitaalinen kirjasto (KDK) -hanke (ks. www.kdk.fi ja www.finna.fi), jossa kehitetään yhteisiä palveluja toi-

mintamalleja. KDK tarjoaa Finna-palvelun, jonka kautta muistiorganisaatioiden kokoelmatiedot ovat asiakkaiden käytettävissä keskitetyksi⁵, sekä PAS-ratkaisun sähköisen tiedon pitkäaikaissäilytykseen. KDK:ssa kokonaisarkkitehtuuryö on varsin jäsentynyttä. Hankkeella on kirjallinen kokonaisarkkitehtuuridokumentti, ja arkkitehtuuryötä tehdään kokonaisarkkitehtuuriryhmässä, tietoarkkitehtuuriryhmässä sekä sektorikohtaisissa arkkitehtuuriryhmissä. Lisäksi KDK ylläpitää standardisalkkua. KDK:n puitteissa myös arkistosektorin toimijoilla on oma viitearkkitehtuurinsa (luonnos).

Tiedesektorilla tiedon saatavuutta ja avointa tiedettä edistää opetus- ja kulttuuriministeriön hallinnonalalla toimiva Avoin tiede ja tutkimus -hanke vuosina 2014–2017. Tämän hankkeen puitteissa CSC tarjoaa tiedon avaamiseen liittyviä yhteisiä ratkaisuja (AVAA, Etsin, IDA). Myös ATT-hankkeessa tehdään viitearkkitehtuuryötä. Tutkimuksen viitearkkitehtuuri on tämän vuoden ATT-hankkeen työsuunnitelmassa, ja arkaluonteisen tiedon tutkimuskäytön viitearkkitehtuurin luonnos on valmistunut huhtikuussa 2015.

Kielitieteen alan toimijoita yhdistää Fin-Clarín, joka on Helsingin yliopiston johtama kansallinen kielivarakonsortio (ks. www.kielipankki.fi), jossa ovat mukana yliopistot ja tieteen tietotekniikan keskus CSC. Fin-Clarín on osa eurooppalaista CLARIN-ERIC-konsortiota, joka puolestaan on osa eurooppalaista ESFRI-tiekarttaa (European Strategy Forum on Research Infrastructures). Fin-Clarín tarjoaa yhteiseen käyttöön tekstikorpuksille Korp-palvelua (<http://korp.csc.fi>), av-korpuksille LAT-palvelua (<http://lat.csc.fi>) ja sanastoaineistoille Sanat-wikiä (<http://sanat.csc.fi>).

Huomioon otettavia sidosarkkitehtuureja ovat myös julkisen hallinnon paikkatiedon viitearkkitehtuuri (Julkisen hallinnon paikkatiedon viitearkkitehtuuri) ja perustietovarantojen viitearkkitehtuuri (Perustietovarantojen viitearkkitehtuuri).

Digitoinnin laatukriteereistä on Kansallisarkiston ohje, joka on kuitenkin vanha ja uudistettua versiota odotetaan. Kotuksella on käytössään uudistetun version luonnos.

⁵ Kansallisen näkymän ohella organisaatioilla on mahdollisuus oman näkymäänsä sekä esimerkiksi sektori-kohtaisiin näkymiin.

Keskeiset arkkitehtuuriperiaatteet, jotka koskevat Kotuksen digitaalisia kieliaineistoja, on koottu seuraavaan taulukkoon (taulukko 5).

TAULUKKO 5. Kotuksen digitaalisten kieliaineistojen arkkitehtuuriperiaatteet.

Arkkitehtuuriperiaatteet

Päiväys 1.6.2015

Versio 0.1

Nimi	Perustelu	Periaatteen lähde
Julkisen hallinnon arkkitehtuuriperiaatteita on noudatettava.	Varmistetaan yhteentoimivuus ja kyky hyödyntää yhteisiä ratkaisuja.	JUHTA, KDK
Toimitaan strategialähtöisesti.	Toiminnan tulee olla suunniteltua, jotta voimavarat voidaan kohdentaa oikein.	Kotus
Hyödynnetään yhteisiä ratkaisuja (mm. KDK, ATT, Fin-Clarín).	Ei omia erillisratkaisuja, jos yhteisiä on olemassa (kustannukset, tehokkuus).	Kotus
Tietoturvallisuudesta huolehditaan aineiston koko elinkaaren ajan.	Kaikki tiedot eivät ole julkisia (mm. henkilötietoja).	Kotus, KDK/arkisto
Julkiset aineistot avataan (CC-lisenssi).	Avaaminen lisää vaikuttavuutta ja hyötyjä ja on valtion linjan mukaista.	OKM, Kotus
Aineistojen metatiedoista huolehditaan.	Metatiedot takaavat aineistojen löydettävyyden ja auttavat ylläpidossa.	KDK
Digitaalisen aineiston pitkäaikais säilytys on luotettavaa.	PAS takaa aineistojen säilymisen.	KDK, Kotus
Toiminta on vuorovaikutteista ja osallistavaa.	Halutaan tietää asiakkaiden tarpeet ja hyödyntää joukkoistamista.	Kotus

Kokonaisarkkitehtuurikuvauksen strategiset linjausten lähteenä on ollut Kotuksen edellinen strategia 2012–2015, mutta linjaukset ovat yhdenmukaiset edelleen uuden vuosien 2016–2020 strategian aikana, semminkin kun strategiaa täydentävät aineistotyön visio ja painopisteet vuodelta 2015. Linjaukset on esitetty alla taulukossa (taulukko 6).

TAULUKKO 6. Strategiset linjaukset.

Strategiat: linjaukset

Päiväys 28.4.2015

Versio 0.1

Strategialinjaus	Vaikutukset arkkitehtuuriin	Lähde
kieliaineistoja tarjotaan mahdollisimman laajasti ja vapaasti käytettäväksi	aineistojen avoin lisensointi (JHS 189, Creative Commons); lataus- ja kyselyrajapintaratkaisut	strategia (s. 1)
aineistoja digitoidaan niiden käytettävyyden parantamiseksi	digitoinnin vaatimien prosessien määrittely; tarvittavan tietojärjestelmä- ja palveluympäristön vakiinnuttaminen	kokoelma- ja palvelupolitiikka (s. 16)
digitointi voi perustua varmuuskopioinnin tai sisällön pelastamisen tarpeeseen	ks. edellistä kohtaa; pitkäaikaissäilytysratkaisun käyttöönotto	kokoelma- ja palvelupolitiikka (s. 16)
aineistojen ylläpitäminen ja kehittäminen tapahtuu yhteistyössä muiden toimijoiden kanssa	toiminta konsortioissa ja yhteistyöverkoitoissa	strategia (s. 6)
aineistojen metatiedot saatetaan kattavasti verkkoon	yhteisten metatietopalvelujen käyttöönotto (erit. Finna, Etsin); metatietojen taustajärjestelmä + haravointirajapinta	kokoelma- ja palvelupolitiikka (s. 3)
tietojärjestelmissä, joissa käsitellään arkaluontoisia henkilötietoja, noudatetaan korotetun tietoturvatason vaatimuksia	tietoturvakäytäntöjen vaatimustenmukaisuus + tekninen tietoturvallisuus (salaukset, tunnistautumiset yms.)	kokoelma- ja palvelupolitiikka (s. 3)

Strategiassa sekä kokoelma- ja palvelupolitiikassa on asetettu useita tavoitteita, jotka eivät ole toistaiseksi toteutuneet. Keskeisimmät on koottu seuraavaan taulukkoon (taulukko 7).

TAULUKKO 7. Strategiset tavoitteet.

Strategiat: tavoitteet

Päiväys 28.4.2015

Versio 0.1

Strategialinjaus	Vaikutukset arkkitehtuuriin	Lähde
aineistojen keräämisessä ja kehittämisessä tukeudutaan tutkijayhteisöön ja kansalaisiin etsimällä nykyaikaisia talkoistamisen muotoja	verkostoituminen; talkoistamistarpeen huomioon ottaminen järjestelmissä (vaikuttaminen yhteisten ratkaisujen suunnitteluun)	strategia (s. 5)
toiminta on vuorovaikutteista ja hyödyntää sosiaalista mediaa	vuorovaikutteisuuden mahdollistavan toiminnallisuuden saaminen järjestelmiin	strategia (s. 9)
edistetään mm. sähköisten luvanvaraisten aineistojen saamista sujuvaan verkkokäyttöön	kieliaineistojen keskittäminen yhteisiin ratkaisuihin, joissa käytössä HAKA (+ tarv. muut); ei omia erillISRatkaisuja	kokoelma- ja palvelupolitiikka (s. 4)
omia aineistoja kehitetään ja kartutetaan tukemaan erityisesti kielenhuollon ja sanakirjatyön tavoitteita	kieliaineistojen tietovarantojen integroiminen kielenhuollon ja sanakirjatyön tuotantjärjestelmiin sekä sähköisiin palveluihin	strategia (s. 9)
aineistotyössä otetaan huomioon kokonaisarkkitehtuurinäkökulma sekä organisaation mittakaavassa että kansallisella tasolla luotaessa yhteistä tutkimusinfrastruktuuria	kokonaisarkkitehtuuri kuvattava ja kehittämiskohteet tunnistettava	kokoelma- ja palvelupolitiikka (s. 4)

Kotuksessa ei ole tehty systemaattisesti kokonaisarkkitehtuurityötä, mutta asian tarpeellisuus on tiedostettu ja tilannetta on tarkasteltu ohjaavan tiedon näkökulmasta. Käytännön tasolla Kotuksessa on toteutettu toimenpiteitä, jotka tukevat kokonaisarkkitehtuurityötä, mutta esimerkiksi prosesseja ja palveluja ei ole kuvattu systemaattisesti ja kattavasti. Pääasiallisella Kotuksen kokonaisarkkitehtuurityö on sillä alueella, joka kuuluu tietoturvallisuuden piiriin.

Kotuksessa kokonaisarkkitehtuurille ei ole nimetty vastuutahoa, mutta tiedonhallintaryhmän tehtäviin kuuluu mm. ”tietovarantojen ja tietojen yhteentoimivuuden kehittäminen (metatietoarkkitehtuuri)” sekä ”tiedon linkkaaren hallinnan kehittäminen”. Aineistoja kirjastoyksikön tehtävänä puolestaan on työjärjestyksen mukaan mm. ”koordinoida Kotuksen tietoarkkitehtuurin kehittämistä”.

5 AINEISTOIHIN LIITTYVÄÄ JURIDIikkaA

Kotuksesta on valtion laitoksena säädetty lailla, ja toimintaa ohjaavat myös monet muut lait. Siihen, miten digitaalisia kieliaineistoja voidaan antaa käyttöön, vaikuttavat erityisesti henkilösuoja- ja tekijänoikeuslainsäädäntö. Lainsäädäntö jättää tulkinnanvaraa eikä ota riittävästi huomioon arkisto- ja tutkimusaineistoihin liittyviä erityiskysymyksiä. (Salokannel 2012, 2013.)

5.1 Lainsäädäntöperusta

Kotuksen toimintaa ja näin myös aineistotyötä ohjaavat laki Kotimaisten kielten keskuksista (1403/2011), valtioneuvoston asetus Kotimaisten kielten keskuksista (1566/2011) sekä opetus- ja kulttuuriministeriön asetus Kotimaisten kielten keskuksen suoritteiden maksuista (1209/2016).

Kotus kuuluu arkistolain (831/1994) soveltamisalaan, mutta arkistolakia sovelletaan Kotuksen osalta vain ns. viranomaisasiakirjoihin. Kieliaineistojen arkistoinnista Kotus vastaa erityistehtävänsä perusteella itsenäisesti. Kotusta voi pitää arkistolain perusteluissa tarkoitettuna erityisviranomaisena, koska nykyisen Kotusta koskevan lain esitöissä todetaan, että ”tehtävien hoitaminen edellyttää kieliaineistojen ja arkistojen ylläpitoa ja kehittämistä”. Lisäksi aineistot ovat olleet keskeisessä roolissa jo Kotusta perustettaessa ja erikseen mainittu Kotusta koskevissa säädöksissä 31.12.2011 asti.

Asiakirja- ja tallennejulkisuuden periaate on perustuslaissa ja julkisuuslaissa. Julkisuuslain (621/1999) mukaan viranomaisen asiakirjat ovat ”julkisia, jollei tässä tai muussa laissa erikseen toisin säädetä”. Salassa pidettäviä kuitenkin ovat ”asiakirjat, jotka on vapaaehtoisesti annettu viranomaiselle tutkimusta tai tilastointia varten”. Suurinta osaa kieliaineistoista ei kuitenkaan ole syytä pitää salassa pidettävinä tutkimusaineistoina, sillä kieliaineistojen ylläpitämiseen liittyy tutkimuksen tai tilastotoiminnan lisäksi muita Kotuksen toimintaan liittyviä sekä yleisiä kulttuurisia ja sivistyksellisiä tarkoituksia ja aiemmin on asetuksessakin todettu kokoelmien julkinen perusluonne.

5.2 Tietosuojasta

Henkilötietolakia (523/1999) sovelletaan henkilötietojen käsittelyyn silloin, kun tiedot muodostavat henkilörekisterin tai sen osan. Laki koskee automaattisen käsittelyn lisäksi myös muunlaista käsittelyä, esimerkiksi paperisten arkistokorttien muodostamaa henkilörekisteriä.

Henkilötiedolla tarkoitetaan henkilötietolain mukaan kaikenlaisia luonnollista henkilö tai hänen ominaisuuksiaan tai elinolosuhteita kuvaavia merkintöjä, jotka voidaan tunnistaa häntä tai hänen perhettään tai hänen kanssaan yhteisessä taloudessa eläviä koskeviksi. Henkilötieto kattaa siis hyvin laajasti kaikenlaiset ihmistä koskevat tiedot. Arkaluontoisia henkilötietoja ovat mm. henkilön yhteiskunnallista, poliittista tai uskonnollista vakaumusta, etnistä taustaa, rikollista tekoa, rangaistusta tai muuta rikoksen seuraamusta, henkilön terveydentilaa, sairautta tai vammaisuutta kuvaavat tiedot.

Luonnollisella henkilöllä tarkoitetaan yleensä elossa olevaa henkilö, mutta henkilötietolakia sovelletaan myös kuolleisiin kuitenkin niin, että yksityisyyden suojan tarve vähenee ajan myötä. Huomattava on kuitenkin se, että kuolleen ihmisen henkilötiedot voivat välillisesti kertoa myös elossa olevista henkilöistä, esimerkiksi heidän lapsuuden elinoloistaan.

Henkilötietolain mukaan henkilötietojen käsittelyn tulee olla asiallisesti perusteltua rekisterinpitäjän toiminnan kannalta. Samoin henkilötietojen käsittelyn tarkoitukset sekä se, mistä henkilötiedot säännönmukaisesti hankitaan ja mihin niitä säännönmukaisesti luovutetaan, on määriteltävä ennen henkilötietojen keräämistä tai muodostamista henkilörekisteriksi. Henkilötietoja saa käsitellä rekisteröidyn yksiselitteisesti antamalla suostumuksella ja muun muassa silloin, jos käsittely johtuu rekisterinpitäjälle laissa säädetystä tai sen nojalla määrätystä tehtävästä tai velvoitteesta. Arkaluontoisten henkilötietojen käsittely on sallittu rekisteröidyn nimenomaisella suostumuksella tai muun muassa silloin, kun käsittely johtuu välittömästi rekisterinpitäjälle laissa säädetystä tehtävästä tai kun tietoja käsitellään historiallista tai tieteellistä tutkimusta taikka tilastointia varten.

Kotuksella on oikeus kieliaineistojen osalta henkilötietojen, myös arkaluontoisten, käsittelyyn lain nojalla määräytyvän tehtävän perusteella. Yksityisyyden suojaan liittyvät ky-

symykset arvioidaan jo keruuhankkeita suunniteltaessa, ja tavoitteena on, että arkaluontoisia henkilötietoja tallennetaan vain siinä määrin kuin on välttämätöntä. Keruun kohteena olevien henkilöiden kanssa tehdään kirjalliset sopimukset, joissa määritellään aineiston käyttötarkoitus. Välittömien tutkimus- ja muiden tarpeiden lisäksi kiinnitetään huomiota aineiston jatkokäyttöön ja arkistointiin, ja pyritään löytämään intressien tasapaino yksityisyyden suojan ja käyttömahdollisuuksien välillä. Henkilörekistereistä laaditaan lainmukaiset rekisteriselosteet. Tietojärjestelmissä, joissa käsitellään arkaluontoisia henkilötietoja, noudatetaan korotetun tietoturvatason vaatimuksia.

Kotuksen salassa pidettävien kieliaineistojen tutkimuskäytössä sovelletaan julkisuuslain 28 §, joka sallii ”yksittäistapauksessa luvan tietojen saamiseen salassa pidettävästä asiakirjastaan tieteellistä tutkimusta, tilastointia taikka viranomaisen suunnittelu- tai selvitystyötä varten, jos on ilmeistä, ettei tiedon antaminen loukkaa niitä etuja, joiden suojaksi salassapitovelvollisuus on säädetty”. Tällöinkään ”lupaa ei saa antaa vastoin suostumuksessa tiedon käytölle ja luovutukselle asetettuja ehtoja”. Suostumuksen voi tulkita niin, että jos se, jonka suojaksi salassapitovelvollisuus on säädetty, luovuttaa aineiston Kotuksen kieliaineistojen osaksi vapaaehtoisesti ilman erityisiä ehtoja, hän antaa suostumuksena aineiston kaikkeen Kotuksen harkinnan mukaisesti hyväksyttävään käyttöön. Ennen 1.12.1999 kerättyjen aineistojen osalta julkisuuslain voimaantulosäännös (37 §) helpottaa tätä arviointia, kun ”Ennen lain voimaantuloa asianomaisen suostumuksella kerättyjä tietoja saa käyttää ja luovuttaa tieteellistä tutkimusta varten siitä riippumatta, mitä 28 §:ssä on säädetty, jos on ilmeistä, että tietojen tällainen käyttö ja luovutus ei olennaisesti poikkea niistä tarkoituksista, joita varten tiedot on annettu.”

Salassa pidettävän aineiston tutkimuskäyttöön antamisesta päättäminen sisältää julkisuuslain esitöiden mukaan sekä oikeusharkintaa että tarkoituksenmukaisuusharkintaa niin, että pääpaino on oikeusharkinnalla ja että tarkoituksenmukaisuuden perusteella luvan epääminen ”voinee tulla vain harvoin kysymykseen”. Lisäksi todetaan, että ”Jos viranomaisen ei anna lupaa, sen on tehtävä asiassa päätös. Viranomaisen päätökseen saa hakea muutosta samassa järjestyksessä kuin muutoinkin lakiehdotuksen mukaisissa asioissa, viime kädessä valittaa korkeimpaan hallinto-oikeuteen.”

Henkilörekisteri, joka ei ole enää rekisterinpitäjän kannalta tarpeellinen, on hävitettävä, ellei talletettuja tietoja ole erikseen säädetty tai määrätty säilytettäväksi. Lisäksi henkilötietolain 35 § mukaan henkilörekisteri, joka on tieteellisen tutkimuksen kannalta tai

muusta syystä merkityksellinen, voidaan siirtää korkeakoulun taikka tutkimustyötä lakisääteisenä tehtävänä suorittavan laitoksen tai viranomaisen arkistoon, jos kansallisarkisto on antanut siihen luvan pyydettyään asiasta tietosuojavaltuutetun lausunnon.

Silloin kun metatieto sisältää henkilötietoa, sovelletaan henkilötietolakia. Metatietojen näyttämisen lähtökohtana on se, että ensisijaisesti asianomaiselta itseltään tai hänen omaisiltaan tai oikeudenomistajiltaan hankitaan lupa tietojen näyttämiseen. Aina tämä ei ole kuitenkaan käytännössä mahdollista, esimerkiksi kun aineisto on kerätty vuosikymmeniä sitten ja henkilöt eivät ole enää arkiston tunnistettavissa. Tämän vuoksi Arkistosektorin KDK-yhteistyöverkosto pyysi Tietosuojavaltuutetun toimiston lausuntoa suositusluonnoksesta *Henkilötietojen näyttäminen metatietoina internetissä*. Luonnoksessa ehdotettiin, että henkilötietoja voisi näyttää metatietoina rajatusti. Näyttäminen perustuisi sen arvioimiseen, a) onko henkilötietoja aiemmin suostumusperusteisesti näytetty julkisesti, b) ovatko tiedot luotettavia ja c) miten tunnettuja tiedot ovat.

Suositusluonnoksessa harkinnanvaraista tietojen näyttämistä perusteltiin mm. sillä, että ”lähes kaikkien kuolleiden henkilöiden henkilötiedoista *nimi, syntymäaika ja kuolinaika* ja mahdollisesti myös *syntymäpaikka, kuolinpaikka ja ammatti* ovat olleet julkisesti nähtävillä, tyypillisesti julkaistuna kuolinilmoituksina sanomalehdissä, ja ovat edelleen nähtävillä, tyypillisesti esim. hautakivissä tai sanomalehtien digitaalijulkaisujen kuolinilmoituksissa”. Helsingin hallinto-oikeuden 2010 tekemän ratkaisun mukaan hautamuistomerkkejä koskevien kuvien julkaiseminen internetissä ei ole tietosuojan vastaista, sillä muistomerkkien ”tiedot on nimenomaan tarkoitettukin kenen tahansa nähtäväksi paikalla, johon yleisöllä on hautausmaiden tai vastaavien aukioloaikojen puitteissa vapaa ja esteetön pääsy”.

Tietosuojavaltuutettu kuitenkin tyrmää ehdotetun käytännön tarpeista lähtevän arkistosektorin ehdotuksen lausunnossaan vuonna 2015. Hän pitää erityisen ongelmallisena sitä, ”että arkistoyhteistyössä luotaisiin näin omat, erilliset, voimassaolevalle vaihtoehdotiset säännöt henkilötietojen verkkoon viemisestä”.

Tietosuojavaltuutettu toteaa, ”ettei henkilötietoja tulisi luovuttaa avoimessa verkossa ilman rekisteröidyn suostumusta tai nimenomaista lain säännöstä”. Tietosuojavaltuutettu perustelee näkemystään perusoikeusnäkökulmalla, ja toteaa, että ”Suojeltava oikeushyvä

tässä on henkilön itsemääräämisoikeus häntä itseään koskeviin tietoihin”. Tietosuojavaltuutetun mukaan ”Henkilötietojen esittäminen avoimessa verkossa vaikuttaa keskeisesti henkilön tiedolliseen itsemääräämisoikeuteen ja myös hänen yksityisyyteensä”. Tietosuojavaltuutettu huomauttaa, että hakukoneita käyttämällä ihmisistä voi helposti luoda yksityiskohtaisia profiileja.

Tietosuojavaltuutetun lausunnon noudattaminen tarkoittaa sitä, että arkistojen kokoelmista ei monessa tapauksessa ole luotavissa samanlaisia tietokantoja kuin on esimerkiksi kirjastojen kokoelmista. On syytä huomata, että arkistojen suositusluonnoksessa verkossa julkaistavaksi on esitetty senkaltaisia tietoja, jotka ihmisistä yleisesti ovat muutenkin tiedossa ja saatavilla, kuten henkilön nimi, syntymäaika, kuolinaika ja ammatti.

Lausunnossaan tietosuojavaltuutettu suosittaa, että viranomaisen verkkopalvelut toteutettaisiin siten, että julkisuuslaissa tarkoitettu tietojen antaminen toteutettaisiin menettelyllä, jossa toteutuvat vaiheet *pyyntö*, *harkinta* ja *luovutus*. Esimerkiksi terveydenhuollon ammattirekisterin (Julki-TERHIKKI), oikeudenkäyntiavustajaluettelon ja eläinlääkäriluettelon verkkoversiot toimivat niin, että rekisteri ei ole kokonaan verkosta ladattavissa vaan kerrallaan voi hakea yhden henkilön tiedot.

EU:n yleinen tietosuojasetus (GDPR) tuli voimaan toukokuussa 2016 ja sen soveltaminen alkaa 25.5.2018. Asetus korvaa kansallisen henkilötietolain. Kesällä 2017 Oikeusministeriön asettama työryhmä ehdotti, että säädettään uusi tietosuojalaki, joka täsmentäisi ja täydentäisi EU:n yleistä tietosuojasetusta. Parhailtaan monet organisaatiot selvittävät, miten EU:n asetukset ja kansallinen uudistuva lainsäädäntö tulevat käytännössä vaikuttamaan.

5.3 Tekijänoikeudesta

Tekijänoikeuslain (404/1961) mukaan ”sillä, joka on luonut kirjallisen tai taiteellisen teoksen, on tekijänoikeus teokseen, olkoonpa se kaunokirjallinen tahi selittävä kirjallinen tai suullinen esitys, sävellys- tai näyttämöteos, elokuvateos, valokuvateos tai muu kuvataiteen teos, rakennustaiteen, taidekäsityön tai taideteollisuuden tuote taikka ilmetköönpä se muulla tavalla”. Lisäksi ”kirjallisena teoksena pidetään myös karttaa sekä muuta selitettävää piirustusta tai graafista taikka plastillisesti muotoiltua teosta sekä tietokoneohjelmaa”. Tekijänoikeussuoja syntyy, kun henkisen luomistyön tuote ylittää teostasoon eli on

itsenäinen ja omaperäinen. Tekijänoikeuden kesto on 70 vuotta tekijän kuolinvuoden päättymisestä.

Tekijänoikeus antaa yksinoikeuden määrätä teoskappaleiden valmistamisesta ja saattamisesta yleisön saataville muuttamattomana tai muutettuna, käännöksenä tai muunnelmana, toisessa kirjallisuus- tai taidelajissa tai toista tekotapaa käyttäen. Yleisön saataville saattaminen tarkoittaa välittämistä yleisölle johtimitse tai johtimitta, esittämistä julkisesti läsnä olevalle yleisölle, kappaleen tarjoamista levitettäväksi yleisön keskuuteen ja julkista näyttämistä. Tekijänoikeus ei rajoita teoksen arkistointia, mutta vaikuttaa olennaisesti siihen, miten tekijänoikeuden alaista arkistossa olevaa aineistoa voidaan käyttää.

Tekijänoikeuslaki ja -asetus sallivat tiettyjen arkistojen, museoiden ja kirjastojen – Kotus kuuluu tähän joukkoon – valmistaa kappaleita omista kokoelmissaan olevista teoksista muun muassa aineiston säilyttämistä ja säilyvyyden turvaamista varten, mikä mahdollistaa esimerkiksi av-aineistojen digitoinnin. Asiakaskäyttöön voidaan poikkeussäädösten perusteella valmistaa lainakappaleita vahingoittumiselle alttiista julkistetuista teoksista. Lisäksi omista kokoelmissa oleva julkistettu teos voidaan välittää yleisöön kuuluvalla henkilölle tutkimusta tai yksityistä opiskelua varten laitoksen tiloissa yleisölle välittämiseen varatuilla laitteilla. Teos katsotaan julkistetuksi, kun se luvallisesti on saatettu yleisön saataviin. Kotukseen kokoelmien osaksi arkistointia voidaan pitää julkistamisena. Julkistettua teosta voi muun muassa siteerata ja siitä voi valmistaa tai valmistuttaa yksityiseen käyttöön muutamia kappaleita (paitsi sävellys-, elokuvateosta, käyttöesineitä, veistosta tai taideteosta taiteellisin menetelmin), eli asiakkaat voivat valokopioida tai digitoida kokoelmiin kuuluvia teoksia yksityiseen käyttöönsä. Opetus, työ tai ansiotoiminta eivät kuitenkaan ole yksityistä käyttöä toisin kuin esimerkiksi opiskelu. Silloin kun Kotuksen kieliaineistoihin kuuluvan teosta halutaan käyttää tekijänoikeudenhaltijan lupaa edellyttävällä tavalla eikä tekijänoikeudenhaltijan kanssa ole tehty sopimusta hyväksytyistä käyttötarkoituksista esimerkiksi arkistoinnin yhteydessä, asia selvitetään tapauskohtaisesti.

Arkistoissa on usein aineistoja vuosikymmenten takaa ja hyvin suurelta määrältä eri henkilöitä. Lisäksi aineistokokonaisuuden sisällä voi olla vaihtelevan laatuista tietoa, jolloin on käytännössä mahdotonta tapaus tapaukselta tulkita, mikä on tekijänoikeuden suojaama ja mikä ei. Samoin tekijöiden tai heidän jälkeläistensä tavoittaminen voi olla mahdotonta.

Suurin osa Kotuksen kieliaineistoista on tekijänoikeudesta vapaita siksi, että teoskynnys ei ylity, että tekijänoikeuden suoja-aika on umpeutunut tai että on kyse tekijänoikeuslaissa mainituista vapaista teoksista. Tekijänoikeuden lähioikeudet voivat kuitenkin vaikuttaa. Erityisesti valokuvien osalta on huomattava, että tavanomaiset teostason alapuolelle jäävät valokuvat saavat 50 vuoden suojan kuvan ottohetkestä alkaen, mutta ennen vuotta 1966 julkistetut tavalliset valokuvat jäävät suojan ulkopuolelle. Tietokannat puolestaan voivat saada lähioikeudellista suojaa joko teoksina, tietokantoina *sui generis* tai luetteloina. Tietokantoihin liittyvät kysymykset ovat olennaisia erityisesti silloin, kun yhteishankkeissa luodaan uusia tietokantamuotoisia kieliaineistoja. Oikeuksista on sovittava tällöin kirjallisesti.

Uusissa aineistohankkeissa tekijänoikeudet selvitetään jo suunnitteluvaiheessa. Tekijänoikeudet ja niiden lähioikeudet tunnistetaan, ja tekijöiltä hankitaan sopimuksin joko täydet tai käyttötärpeisiin nähden riittävän laajat ns. taloudelliset tekijänoikeudet.

6 NIMIARKISTON DIGITALISOINTI

Nimiarkiston digitalisoinnin tarkasteleminen (2014–2017) kokonaisuudessaan luo hyvän kuvan sitä, kuinka pitkä matka alkuperäisdatasta voi olla käyttäjälle näkyvään valmiiseen palveluun. Tämä digitalisointihanke on ollut osa kehittämistehtävääni. Olen toiminut hankkeen projektipäällikkönä.

6.1 Nimiarkisto

Nimiarkisto on keskeinen osa Kotuksen kansallisesti merkittävää kielitieteen tutkimusinfrastruktuuria ja kansallista kulttuuriperintöä. Nimiarkisto kokoaa ja ylläpitää suomen- ja saamenkielisiä nimiaineistoja tutkimuksen ja nimistönhuollon tarpeisiin.

Nimiarkistossa on noin 2,7 miljoonan nimilipun verran paikannimitietoa. Nimilippujen lisäksi tietoa on sidoksina ja käsikirjoituksina. Nimilipussa on yleiskielistetty nimen haakuasu, nimen murremuoto ja taivutus, sijainti (pitäjä, kylä, talo ja karttasignum), paikan laji, mahdollisesti selitys sijainnista sekä kaikki haastatteleamalla saatu nimeä tai paikkaa valaiseva muistitieto. Sanallisten sijaintitietojen lisäksi nimet on paikannettu keruukarttoihin (n. 10 000 kpl) tehdyin merkinnöin. Nimiarkisto on käytetyimpiä Kotuksen arkistoja. Arkistossa on arvokasta tietoa paitsi nimistön- ja kielentutkijoille myös arkeologeille, historianutkijoille, kansatieteilijöille ja luonnontieteilijöille. Myös sukututkijoille arkisto tarjoaa kiinnostavaa oheis- ja taustamateriaalia. Julkishallinnon kannalta Nimiarkisto on tärkeä nimistönhuollon lähde.

Suomen ruotsinkielinen paikannimistö on tallennettu Svenska litteratursällskapetiin. Hanketta on laajennettu vuonna 2016 kattamaan myös ruotsinkielinen nimistö, noin 380 000 nimitietoa ja 1 000 keruukarttaa. Samoin Suomen saamenkielinen paikannimistö on tarkoitus saada osaksi palvelua tulevina vuosina yhteistyössä saamelaisyhteisöjen kanssa.

6.2 Alkuvaiheet

Kotuksen Nimiarkiston digitalisointihanke alkoi 2014, ja tavoitteeksi asetettiin arkiston suomenkielisten paikannimikokoelmien digitalisoiminen vuoden 2017 loppuun mennessä.

Valtiovarainministeriö on selvittänyt tietovarantojen avaamisen rahoitustarpeita vuosittain osana ns. valtionalouden kehysvalmistelua, ja tällöin eri ministeriöiltä on pyydetty omalta hallinnonalaltaan rahoitustarve-esityksiä. Vuonna 2012 opetus- ja kulttuuriministeriö järjesti hallinnonalansa laitoksille asiasta keskustelutilaisuuden. Kotuksen etukäteen laatimassa esityksessä Nimiarkiston digitalisointia ei ollut ehdotettu, mutta keskustelussa esiin noussut mahdollisuus digitoida Kotuksen laajoja arkistoja herätti ministeriön mielenkiinnon, ja ministeriö pyysi Nimiarkiston digitalisoinnista alustavan suunnitelman. Vaikka Kotus ei ollut strategisesti linjannut arkistojen digitointia aiemmin samana vuonna päivitettyssä strategiassaan, kokoelmien digitointiprioriteetit oli esitetty syksyllä 2012 valmistuneessa Kotuksen kieliaineistojen ja kirjaston kokoelma- ja palvelupolitiikassa. Nimiarkiston digitointi oli esitetty ensisijaisena, jos laajamittainen digitointi tulee mahdolliseksi (mts. 16):

”Nimiarkiston aineistoilla on eniten ulkoisia asiakkaita, ja pitäjänkokoelmia on jo jonkin verran digitoitu asiakastilausten muodossa. Nimiarkiston aineistot eivät tule julkaisujen muodossa käyttöön, joten niiden digitointi myös laajassa mitassa on perusteltua.”

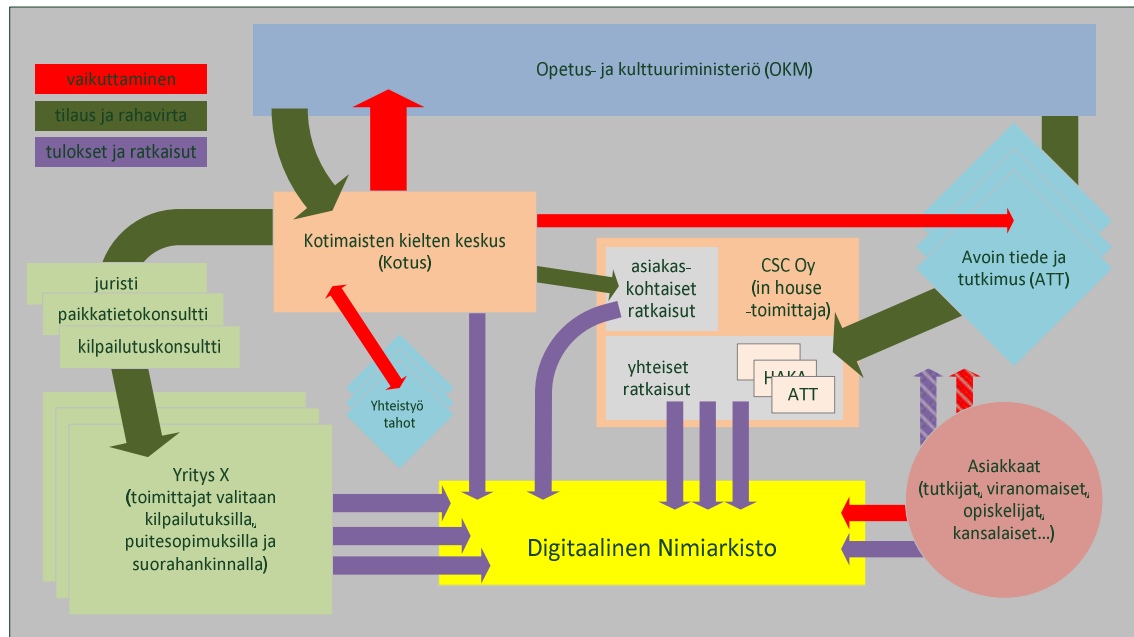
Nimiarkiston digitalisointi päättyi kehusehdotuksena opetus- ja kulttuuriministeriön toiminta- ja taloussuunnitelmaan 2014–2017 (s. 39):

”Kotimaisten kielten keskuksen Nimiarkiston paikkatietokokoelman nimiin liittyvät tiedot avataan osin rajoituksetta (mm. nimet, sijainnit) ja osin rajoitetusti (mm. henkilöjä koskeva muistitieto).”

Rahoitusta hankkeelle ei saatu tietovarantojen avaamiseen tarkoitetuista varoista, mutta vuonna 2014 opetus- ja kulttuuriministeriö myönsi hankkeen aloittamiseen 263 000 euron erityisavustuksen vuosina 2014 ja 2015 käytettäväksi. Kotus on tehnyt päätöksen Nimiarkiston digitalisoinnin aloittamisesta ja suuntaviivoista 13.8.2014.

Vuodeksi 2015 Kotus osoitti hankkeeseen myös muuta rahoitusta niin, että noin kolmannes arvioiduista miljoonan euron (+ arvonlisäverokulut) kokonaiskustannuksista oli katettu. Jatkorahoitusta on saatu vuosittain 2015, 2016 ja 2017 yhteensä 755 000 e. Jonkin verran rahoitusta on saatu myös kunnilta, muilta yhteisöiltä ja säätiöiltä.

Kuva 5 havainnollistaa karkeistetusti toimijoiden välisiä suhteita.



KUVA 5. Nimiarkiston suomenkielisten paikannimikokoelmien digitalisointihankkeen toimijoiden suhteet. (Toni Suutari 2015.)

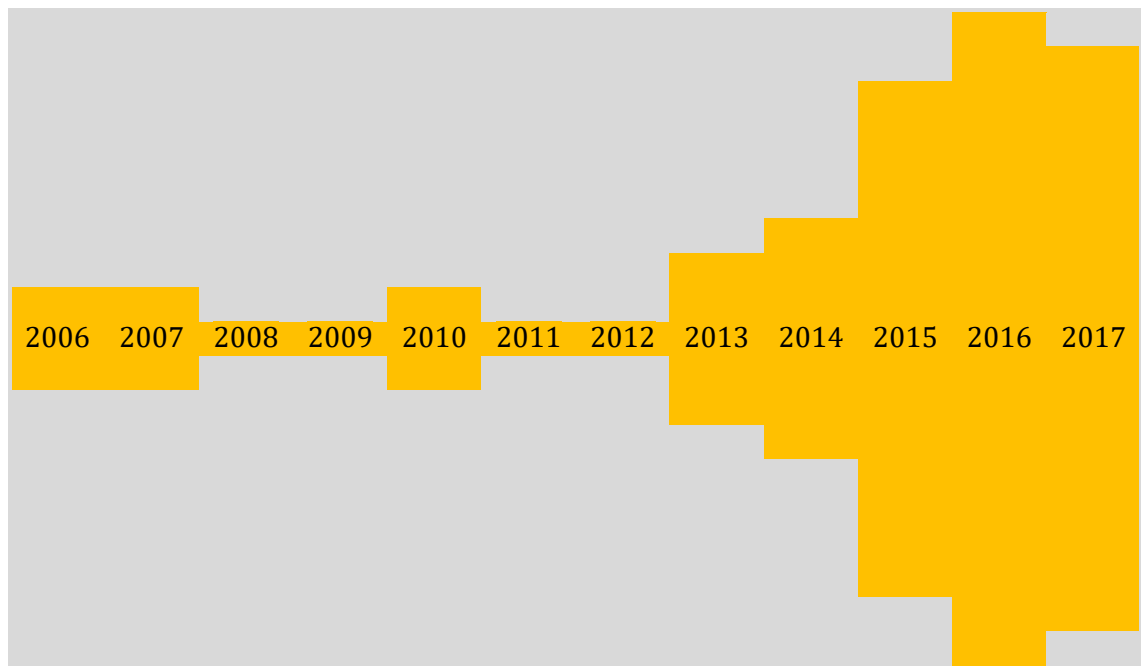
Nimiarkiston kokoelmien metatiedot oli digitoitu sähköiseksi tietokannaksi paperikorkeilta jo aiemmin vuodesta 2010 alkaen. Tämä paransi merkittävästi edellytyksiä nimiarkistoistojen massadigitoinnille. Jonkin verran oli tehty muitakin työtä, joka oli omiaan pohjustamaan isoa hanketta. Esimerkiksi suoraan sähköiseen muotoon tehdyt kokoelmat oli yhdistetty yhtenäiseksi tietokannaksi vuonna 2012. Lisäksi joitakin Nimiarkiston kokoelmia on digitoitu asiakkaille maksupalveluna. Leimallista on ollut, että kaikki työt digitoinnista tietokantoihin ja käyttöliittymiin on tehty omana työnä vuoteen 2012 asti.

Paikannimitietojen esittämisessä tarvittavia tietojärjestelmiä ja ohjelmistoja alettiin systemaattisesti kehittää opetus- ja kulttuuriministeriön erillisrahoituksella vuonna 2013 Avoin data ja paikkatieto -hankkeen myötä, ja tämä työ jatkui vuonna 2014. Avoin data ja paikkatieto -hankkeessa avattiin verkossa kielikartastoja ja muuta paikkatietoaineistoa.

Nimiarkiston digitalisointia edeltävät ja siihen varsinaisesti kuuluvat eri vaiheet ja projektit nivoutuvat kokonaisuudeksi, vaikka tällaista strategista kokonaistavoitetta ei ole ollut. Alla oleva kuviosta (kuvio 1) näkyvät eri vaiheet ja niiden ajoittuminen. Jälkimmäinen kuvio (kuvio 2) puolestaan havainnollistaa eri vaiheiden intensiteettiä resurssien näkökulmasta.

<i>Nimiaineistojen digitalisoinnin edistäminen</i>			
2006	Nimiarkiston kokoelmakortistot tietokannaksi		
2007			
2008			
2009			
2010	Paikannimitietokanta ja aineistojen digitointia		
2011			<i>Nimiarkiston digitalisointi</i>
2012		<i>Avoin data ja paikkatieto</i>	Alustava suunnitelma OKM:lle
2013		Paikkatietodatan käsittely ja perusratkaisut	
	2014	Paikkatietodatan julkaiseminen ja palvelut	Rahoitus (Q3) OKM:ltä ja hankkeen aloitus
	2015		Digitointityöt ja konseptisuunnittelu
		2016	Digitoinnin jatko ja tietojärjestelmien suunnittelu
		2017	Palvelukokonaisuuden valmistuminen

KUVIO 1. Nimiarkiston digitalisointiin liittyviä vaiheita 2006–2017.



KUVIO 2. Nimiarkiston digitalisoinnin liittyvien vaiheiden intensiteetti 2006–2017 oman ja ostopalvelun työmäärien perusteella arvioituna.

Nimiarkiston digitalisointihanke on esillä myös Kotuksen toiminta- ja taloussuunnitelmassa 2016–2020:

”Digitaaliarkistosta kehitetään keskeinen kansallinen tietovaranto, jolla hallitaan Suomen historiallista ja epävirallista, käytössä olevaa paikannimistöä sekä niihin liittyviä paikkoja yksilöivien tunnisteiden avulla. Tunnisteita hyödynnetään yhä enemmän tutkimus- ja muistiorganisaatioiden ylläpitämisen tiedon hallinnassa, ja tähän tarkoitukseen virallisen nykykarttanimistön sisältävä Maanmittauslaitoksen Paikannimirekisteri ei yksin riitä. Paikannimirekisterillä ja Nimiarkistolla on selkeä työnjako ja aineistot täydentävät toisiaan. Yhteistyötahot ja kansalaiset otetaan laajasti mukaan uuden nimistötiedon keräämiseen ja olemassa olevan tiedon yhteisölliseen rikastamiseen Tätä nimistön nykytilaa kuvaavaa tietoa tarvitaan muun muassa virallisen nimistön suunnittelussa ja -huollossa, joten digitaalisen Nimiarkiston kehittäminen epävirallisen käytössä olevan paikannimistön auktoriteettitietokannaksi liittyy keskeisesti Kotuksen ydintehtäviin kuuluviin kiel- ja nimistöhuoltoon.”

6.3 Ostopalvelujen hankinta

Nimiarkiston paikannimiaineiston digitoinnin, digitaalisen aineiston käsittelyn ja digitaalisen palvelukokonaisuuden muodostamisen vaatima työ hankitaan pääosin ostopalveluina. Tietojärjestelmä- ja palveluosuus toteutetaan yhteistyössä Kielipankin ja CSC:n kanssa. Kilpailutuksen valmistelussa hyödynnettiin Hanselin sekä Paikkatietokonsultit Oy:n asiantuntemusta. Kotus julkaisi osana hankintojen valmistelua tietopyynnön HILMA-portaalissa vuoden 2014 lopussa, ja tammikuussa 2015 käytiin yritysten kanssa keskustelua heidän näkemyksistään. Tämän jälkeen edettiin kilpailutuksiin, jotka toteutettiin vuoden 2015 aikana vaiheistetusti.

Nimiarkiston paikannimikokoelmien digitalisoinnin kilpailutus on toteutunut seuraavasti vuoden 2015 aikana, lukuun ottamatta pieniä suorahankintoja, jotka jatkuvat koko hankkeen ajan:

- Kilpailutuksen ennakkotoimet vuonna 2014 ja alkuvuonna 2015
 - ennakkoilmoitus ja tietopyyntö Nimiarkiston paikannimikokoelmien digitalisointiin liittyvistä vuodelle 2015 ajoittuvista kilpailutuksista
- Hanselin puitejärjestelyn mukainen kevennetty kilpailutus
 - osaprojekti 1 (nimilippujen digitointi, n. 2,3 milj. kpl)
 - osaprojekti 2 (sidosten ym. erikoisaineistojen digitointi, n. 0,3 milj. kpl)
- Avoimen menettelyn kilpailutuksella
 - osaprojekti 3 (keruukarttojen digitointi mikrofilmeistä, n. 10 000 kpl)
 - osaprojekti 4 (skannatun nimilippuaineiston tietokantaistaminen)
 - osaprojekti 5 (digitoitujen keruukarttojen georeferointi ja aineiston paikantaminen)
- Hanselin puitejärjestelyllä ja yhteistyössä CSC:n kanssa
 - osaprojekti 6 (tietojärjestelmät), 1. vaiheessa suunnittelu, 2. vaiheessa toteutus
- Lisäksi suorahankintana konsultti- ja asiantuntijapalveluja

Nimiarkiston digitalisointihankkeessa on monta vaihetta ennen kuin valmis digitaalinen data on valmiina sijoitettavaksi tietojärjestelmiin ja ohjelmistoihin. Koska CSC on valtion täysin omistama yhtiö, se on hankintalain kannalta ns. in house -toimittaja, jolta palvelut voidaan ostaa kilpailuttamatta.

6.4 Osallistaminen

Digitaalisessa arkistossa ei ole kyse vain olemassa olevan tiedon käytettävyyden parantamisesta, vaan sähköinen toimintaympäristö tarjoaa tärkeitä mahdollisuuksia tietojen rikastamiseen ja täydentämiseen. Kansalaiset voivat tuottaa tietoa siitä, ovatko arkistoon tallennetut nimet vielä käytössä, onko nimi korvautunut uudella, onko nimettyä paikkaa enää olemassakaan tai mitä kokonaan uutta nimitystä on syntynyt. Samoin voidaan täydentää paikkoihin ja nimiin liittyvää taustatietoa sekä muuntaa kuvamuotoisista lipuista tietokantamuotoon nimiin liittyvää tietoa. Sähköisessä järjestelmässä arkiston alkuperäinen asiantuntijatieto ja kansalaisten tuottama tieto täydentävät toisiaan ja voidaan esittää niin, että käyttäjälle on selvää, milloin on kyse asiantuntijatiedosta ja milloin sitä täydentävästä tiedosta, joka voi olla luonteeltaan varioivampaa. Sähköinen järjestelmä helpottaa myös asiantuntijatyönä kerätyn uuden tiedon käyttöön saattamista.

Nimiarkiston digitalisointiprojektin on tarkoitus olla siinäkin mielessä uudennainen, että käytettävyyteen tullaan kiinnittämään erityistä huomiota, rakennetaanhan kansalaisten käytössä olevaa palvelua. Kehittämisessä hyödynnetään toiminta- ja käyttötarinoita ja käytettävyydestä (Sinkkonen et al. 2006: 29–, 275–).

6.5 Palvelut asiakkaille

Alun perin tarkoituksena oli, että digitaalisen Nimiarkiston tarvitsemat palvelut toteuttaa CSC osana Avoin tiede ja tutkimus -hankkeen tarjoaman AVAA-palvelua, jonka kautta Kotuksen Avoin data ja paikkatieto -hankkeen tuottamat aineistot on julkaistu. Kotus keskusteli hankkeen tietojärjestelmätarpeista 12.1.2017 OKM:n, CSC:n ja Kielipankin kanssa. Todettiin, että nimiaineistojen käyttöön saattamista aletaan edistää yhteistyössä Kielipankki-konsortion kanssa. Kielipankilla jo olemassa oleva sana-aineistojen julkaisemiseen kehitetty Sanat-wiki, joka pohjautuu Tieteen termipankin wiki-ratkaisuun, tarjoaa hyvän pohjan digitaalisen paikannimitiedon julkaisemiselle. Digitaalisen Nimiarkiston kannalta keskeistä on kehittää ratkaisuja, jotka mahdollistavat olemassa olevien tietojen rikastamisen ja uuden tiedon keräämisen kansalaisia laajasti osallistamalla.

Wikijulkaisu tuotetaan käyttäen Wikibase-laajennosta MediaWikissä (sanat.csc.fi). Tietokantamuotoinen aineisto sijoitetaan Wikibasea varten MariaDB-kantaan. Tarvittaessa käytetään myös PostGIS-kantaa paikkatiedon käsittelyä varten (mm. aluekohteet).

Rakenteellinen data on datawikissä (Wikibase), ja se esitetään osana wiki-artikkelia Paikannimiwikissä. Ratkaisu on samanlainen kuin Wikipedian ja Wikidatan keskinäisessä kommunikoinnissa. Käyttäjänhallinnassa hyödynnetään eDuunin ID-ratkaisua, joka tarjoaa HAKA- ja Virtu-verkoston lisäksi mahdollisuuden käyttää muita, kansalaisten ulottuvilla olevia tunnuksia.

Geoserverin kautta tarjotan tietokantamuotoisen aineiston kysely- ja latausrajapintapalvelut sekä kartta-aineiston taustakartta-palvelu (tiilitys).

6.6 SWOT

Nimiarkiston digitalisointihanketta on sen alkuvaiheissa tarkasteltu SWOT-analyysin kautta (kuvio 3).

sisäinen ympäristö	S vahvuudet - hankkeen tavoite, digitaalinen Nimiarkisto selkeä kokonaisuus - henkilökunnan sitoutuneisuus ja substanssiosaaminen - kokemus digitoinnista ja digitaalisista aineistopalveluista	W heikkoudet - niukat resurssit, tästä johtuva haavoittuvuus - vähän kokemusta palvelujen ostamisesta (itse tekemisen perinne) - projektikulttuurin puutteet - strategia- ja muutosjohtamisen heikkoudet
	O mahdollisuudet - kiinnostusta herättävä tavoite - eri käyttäjäryhmät ja yhteistyökumppanit tukena - tietovarantojen avaaminen yleisten tavoitteiden mukaista (rahoitusta saatu) - kansalaisten kiinnostus ja joukko-voima - digitoinnin ja datan avaamisen tuomat uudet mahdollisuudet	T uhat - rahoituksen loppuminen kesken - kilpailutuksista ja palveluntoimitajista johtuvat riskit - mutkikas kokonaisuus, jossa kaikkien ei voi itse vaikuttaa - ei kyetä vastaamaan (kaikkien) käyttäjien odotuksiin ja tarpeisiin, maineriski - onnistutanko osallistamaan ja joukkoistamaan

KUVIO 3. SWOT-analyysi Nimiarkiston digitalisoinnin alkuvaiheesta vuodelta 2015. (Toni Suutari.)

Hanke on nyt loppusuoralla, joten voidaan todeta, että analyysi oli monin kohdin oikeaan osuva. Avaan seuraavissa alakohdissa analyysin eri kohtia tämänhetkisen tilanteen valossa.

Vahvuudet

Nimiarkiston digitalisoinnissa kohteena on laadukas kokonaisuus, johon liittyy vahva substanssiosaaminen, kuten tietämys kokoelmien sisällöstä ja käyttömahdollisuuksista. Henkilökunta on sitoutunutta. Organisaatiolla on kohtalaisesti kokemusta kieliaineistojen digitoinnista ja digitaalisten palveluiden rakentamisesta.

Heikkoudet

Kotuksen kieliaineisto-osaamisen resurssit ovat hyvin vähäiset, mistä seuraa se, että yhdenkin avainhenkilön poistuminen voi vaarantaa hankkeen. Lisäksi näin laajasta ostopalveluna toteutettavasta kokonaisuudesta ei ollut aiempaa kokemusta, esimerkiksi kokemusta hankintojen kilpailuttamisesta yms. ja it-alan ostopalveluista on organisaatiossa melko vähän. Niin ikään organisaation projektikulttuuri ei ole ollut vakiintunutta, esimerkiksi ajantasainen projektikäsikirja otettiin käyttöön vasta vuonna 2016. Nimiarkiston digitalisointihankkeen alku kärsi resurssipulasta ja siitä, ettei ollut valmiina selkeitä ja vakiintuneita projektikäytäntöjä sekä välineitä, joiden avulla resursseja, aikatauluja ja tavoitteiden toteutumista tehokkaasti olisi seurattu. Pahimmillaan organisaation ja projektien johtamisen ongelmat olisivat voineet heijastua hankkeeseen niin, että innovatiivisuus ja tuloksellisuus kärsivät. Olisi kyettävä keskittymään suorituksen johtamista (ks. Järvinen et al. 2014).

Mahdollisuudet

Digitoitava aineisto on laajasti kiinnostusta herättävää ja tarjoaa digitaalisena avoimena datana monia uusia käyttömahdollisuuksia. Digitaalinen arkisto parantaa tutkimusedellytyksiä, hyödyttää viranomaistoimintaa ja avaa kulttuuriperintöä kaikkien käyttöön. Yhteistyökumppaneilta ja asiakkailta on saatavissa tukea, ja julkisen sektorin toiminnan hyvänä puolena on se, ettei synny kilpailuasetelmaa. Joukkoistaminen ja osallistaminen ovat tärkeitä mahdollisuuksia, joiden avulla voidaan kartuttaa ja jalostaa aineistoa. Ulkoisen ympäristön luomana mahdollisuutena voi pitää myös rahoitushalukkuutta, jonka taustalla on julkisten tietovarantojen avaamisen edistäminen. Laaja hanke antaa Kotukselle mahdollisuuden näyttää rahoittajalle ja asiakkaille kyvykkyytensä paitsi hankkeen teknisessä

toteuttamisessa myös siinä, että Kotus voi toimia innovatiivisena suunnannäyttäjänä (esim. kansalaisten osallistamisessa).

Uhat

Ulkoisena uhkana on ollut hankkeen rahoituksen jatkuminen, varsinkin kun yleinen taloudellinen tilanne on ollut heikko. Rahoitus on kuitenkin nyt olemassa vuoden 2017 loppuun saakka. Kuitenkin mahdolliset ja toivottavat jatkohankkeet ovat edelleen riippuvaisia julkisen sektorin erillisrahoituksesta. Organisaation budjettivaroin suuren hankkeen toteutuminen olisi mahdotonta tai erittäin hidasta. Käytännöllisiä uhkia voi aiheutua palvelujen hankinnasta (mm. heikko palveluntoimittaja tai kilpailutuksesta tehtävä valitus markkinaoikeuteen). Lisäksi kokonaisuus on mutkikas hallittavaksi, ja siihen liittyy monia kohtia, joissa ei toimita tilaaja–tuottaja-mallin mukaisesti, eli jossakin määrin toimintaympäristöstä voi aiheutua aikataulullisia tai sisällöllisiä rajoituksia sille, mikä on mahdollista toteuttaa. Koko kansan palvelua suunniteltaessa uhkana on myös se, että käyttäjien tarpeita ja odotuksia ei kyetä tunnistamaan tai niihin vastaamaan. Erityisesti käyttäjien osallistaminen voi kariutua, jos palvelu ei vakuuta käyttäjiä.

6.7 Nimilippujen henkilötiedot

Tietokantamuotoon siirrettävä nimilippujen keskeisin tietosisältö sekä pistekoordinaatti-paikannukset voidaan avata CC-lisensioituna avoimena datana. Nämä tiedot eivät sisällä henkilötietoja, jotka estäisivät aineiston julkaisemisen Internetissä. Sen sijaan nimilippujen skannattujen kuvien saattaminen verkon kautta asiakkaiden käyttöön on ongelmallisempaa, koska kuvat sisältävät kaiken tiedon, myös henkilötiedot muun sisällön lomassa. Tämän vuoksi hanke on tilannut asianajotoimistolta tietosuojaan liittyvän selvityksen.

Nimiarkiston aineistoa on kerätty noin sadan vuoden ajan, ja erillisiä suostumuksia ei ole kysytty. Tämä olisikin ollut käytännössä mahdotonta, sillä keruumetodin vuoksi aineistossa on paljon toisen käden tietoja ja tietoja henkilöistä, jotka jo tietojen keruuhetkellä ovat olleet kuolleita. Suurin osa nimiin liittyvistä henkilötiedoista koskee kuolleita henkilöitä.

Henkilötietolain 14 § mukaan henkilötietoja saa käsitellä tietyin edellytyksin historiallista tai tieteellistä tutkimusta varten, ja nämä edellytykset täyttyvät, mutta tällä perusteella

aineistoa ei voida laittaa vapaasti verkkoon, koska tällöin tietoja voitaisiin käyttää muuhunkin tarkoitukseen.

Asianajotoimiston tulkinta on, että olemassa olevan lainsäädännön perusteella arkiston sisältämiä henkilötietoja ei voi saattaa sähköisesti vapaasti asiakkaiden saataville, ei Internetissä eikä edes esimerkiksi laitoksen tiloissa olevan yleisöpäätteen kautta, vaikka paperimuotoinen arkisto on aina ollut asiakkaiden (myös muiden kuin tutkijoiden) vapaasti käytettävissä. Digitoitujen nimilippujen kuvien osalta ajatuksena on, että nimilippujen kuvat annetaan verkon kautta vahvasti tunnistetuille (HAKA) tutkijakäyttäjille henkilökohtaisen käyttöluvan perusteella, mutta näyttää siltä, että muille kuin tutkijoille aineistoa ei voitaisi antaa sähköisenä käyttöön edes lupaprosessia ja vahvaa sähköistä tunnistautumista käyttäen.

Osa nimilipuissa olevista henkilötiedoista on niin vanhoja, että henkilötietolain poikkeussäännösten perusteella ne ovat rajoituksista vapaita, ja kaikkiin nimilippuihin ei edes liity henkilötietoja. Aineistoa voidaan lajitella keruuhetken perusteella, mutta muuten vajaan kolmen miljoonan yksikön läpikäyminen henkilötietojen näkökulmasta olisi kuitenkin usean henkilötyövuoden laajuinen tehtävä, johon ei ole resursseja. Näin ollen aineistoa on ainakin toistaiseksi käsiteltävä yhtenäisenä kokonaisuutena.

7 TULEVAISUUDEN SKENAARIOT

Kotuksen aineistojen kehittämisessä on suuria mahdollisuuksia, mutta tosiasiasa lienee pikemmin kyse ajassa elämisestä ja välttämättömästä muutoksesta, ei niinkään uusista innovaatioista. Innovaation syntymahdollisuuksia voisi olla esimerkiksi uudenaikaisessa tavassa yhdistää aineistoja ja sanakirjoja, näin uudistaa sanalähtöistä tiedon esittämistä, mutta ”erilaisten ajatusten törmäyttäminen ja monialainen äänteen kuuleminen”, joita innovaatioprosessissa pidetään tärkeinä (Ojasalo et al. 2014: 82), on haaste. Sama koskee ennakkointia (Ojasalo et al. 2014: 90), jossa organisaatio ei ole aina onnistunut. Esimerkiksi suursanakirjojen verkkojulkaisemiseen ei aikanaan haluttu siirtyä oma-aloitteisesti vaan muutos tapahtui opetus- ja kulttuuriministeriön ohjauksella ulkopuolisen arvioinnin jälkeen (Kotimaisten kielten tutkimuskeskuksen tehtävät -muistio 2010).

Toimintaympäristön muutostekijöiden analyysi, tulevaisuuden vision muodostaminen ja kriittisten menestystekijöiden tunnistaminen ovat strategiatasolla toteutettavia asioita, joiden tulisi tukea operatiivista toimintaa. On hyvä arvioida poliittisen ja hallinnollisen ohjauksen merkitys ja selvittää eri sidosryhmien ja organisaation omat odotukset sekä käyttää kaikkea tätä tietoa läpinäkyvästi strategisia valintoja tehtäessä ja perusteltaessa. Strategisen johtamisen heikkous voi johtaa siihen, että ”keskitytään lillukanvarsiin, mutta isot asiat humahtavat vahingossa läpi” tai että ”jokainen päätös tapellaan kerrallaan”, jolloin toiminta on helposti epäjohtonmukaista (Gustafsson & Marniemi 2012: 15).

Riskinä on, että strategian ja laajemmin päätöksenteon pohjana käytetään näennäistä sisäistä ”tietoa”, oikeastaan näkemyksiä ja mielipiteitä, ja lähinnä vain organisaation omaa näkemystä tukevaa ulkopuolista tietoa, jolloin muuttuvan toimintaympäristön asettamat odotukset voivat jäädä tunnistamatta tai tulla aliarvioiduiksi merkitykseltään. Toisaalta on mahdollista, että strategiaan otetaan konkretisoimattomiksi tavoitteiksi monenlaisia muutosta edellyttäviä asioita, jotka eivät kuitenkaan toteudu organisaation tosiasiallisessa toiminnassa. Jos toiminnan tuloksellisuuden seuraaminen ei ole ajan tasalla, tällainen voi jäädä huomiotta, usein ehkä jopa tarkoituksellisesti.

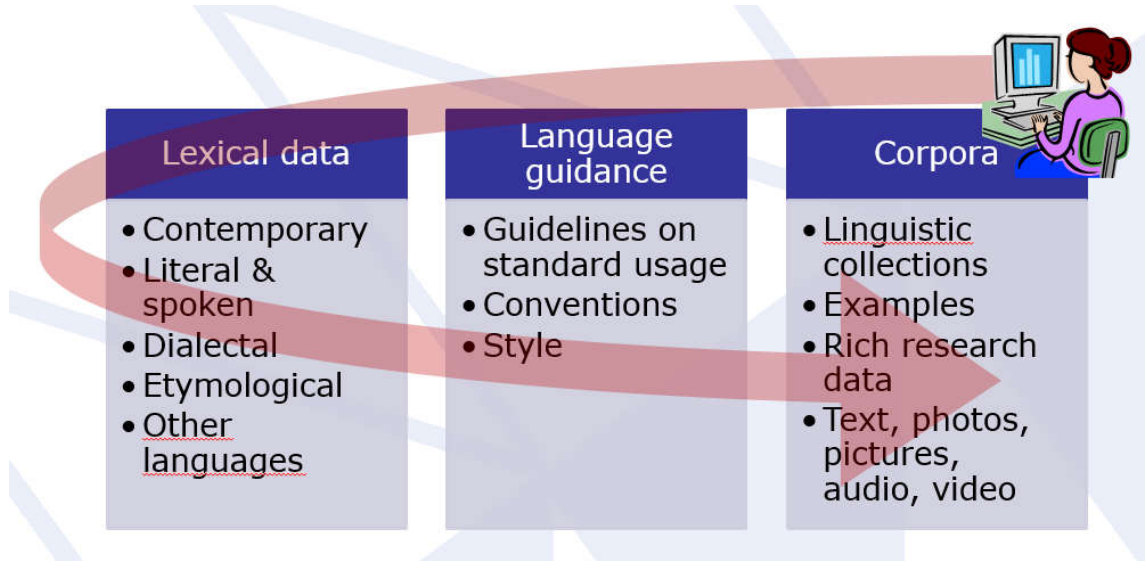
Choon tiedonhallinnan prosessimallissa (Choo 2001) ajatuksena on, että organisaatio kerää ja analysoi tietoa, jotta sen avulla voidaan mukauttaa toimintaa yhteisesti asetettujen

päämäärien ja ulkoisen toimintaympäristön vaatimusten mukaisesti. Strategisen kyvykkyyden kehittämisen strateginen, operatiivinen ja kehityksellinen ulottuvuus edellyttävät tietoa päätöksenteon ja toiminnan ”raaka-aineeksi”. Tiedon hankkiminen ja analysointi korostuvat erityisesti toimintaympäristön muutostekijöiden analyysissä, vaikuttavuus-, asiakas- ja sidosryhmäytyvyyden mittaamisessa ja itsearvioinneissa (Virtanen 2006).

Konkreettisenä esimerkkinä tavoitteena olevasta palvelukokonaisuudesta, konseptista voit mainita Kotuksen suunnitelman aineistot ja sanakirjat yhdistävästä palvelusta. Nyt Kotuksen teksti- ja sanakorpukset, sanakirjojen materiaalina toimivat arkistomateriaalit (esim. murre-sanatiedot) ja eri sanakirjat ovat itsenäisiä kokonaisuuksiaan, jotka kuitenkin olisivat kytkettävissä toisiinsa saumattomaksi kokonaisuudeksi, jossa voisi liikkua sanalähtöisesti sanakirjatiedossa, esimerkkiaineistoissa ja korpuksissa. Tämä edellyttäisi kuitenkin myös toimintakulttuurin laajempaa muutosta. Sanakirjojen toimittaminen perustuu edelleen koko lailla klassiseen sanakirja-ajatteluun, vaikkakin viime vuosina Kotus on siirtynyt julkaisemaan suursanakirjojaan vain verkossa (*Suomen murteiden sanakirja*, *Vanhan kirjasuomen sanakirja*, *Ordbok över Finlands svenska folk mål* ja jo aiemmin sähköisenä julkaistu *Karjalan kielen sanakirja*, sähköisenä ovat olemassa myös *Kielitoimiston sanakirja* ja *Suomen sanojen alkuperä*).

Aineistojen ja sanakirjojen yhdistämisestä (AISA-hanke) on olemassa suunnitelma vuodesta 2013 alkaen. AISA-hankkeeseen on haettu Suomen Akatemian rahoitusta ja hanke on ollut esillä strategisissa tavoitteissa, mutta tilanne on avoinna ja jonkinlainen strateginen linjaus olisi paikallaan. Kuva 6 havainnollistaa hankkeen ideaa, jossa monenlaiset aineistot linkitetään yhdeksi helppokäyttöiseksi kokonaisuudeksi. Yksinkertaisimmillaan sanahaku voisi tuottaa tietoa sanan merkityksestä ja käytöstä eri aikoina, sanaan liittyviä kielenhuollollisia ohjeita sekä viittauksia sanan esiintymiin eri korpuksissa, samoin tilastotietoa.

Aineistojen ja sanakirjojen yhdistäminen toisi merkittävästi lisäarvoa sekä aineistoille että sanakirjoille, mutta se vaatisi merkittäviä resursseja asiantuntijatyöhön. Suuri osa arkistolippumuotoon kerätyistä arkistokokoelmista on sanalähtöisiä (mm. murrearkisto, nyky-suomi, slangit), joten ne olisi kytkettävissä sanakirjatietoon. Toisaalta eri sanakirjat voitaisiin kytkeä toisiinsa, jolloin syntyisi laaja kokonaisuus, joka avaisi sanalähtöisen näkymän, joka ulottuisi etymologiasta murteisiin ja nykykieleen ja jonka kautta laajat arkistokokoelmat olisivat indeksoituna hyödynnettävissä.



KUVA 6. Aineistot ja sanakirjat yhdistävä palvelu. (Toni Suutari, kuva laadittu 2013 Suomen Akatemian ulkoisille arvioijille suunnattuun esitykseen.)

Aineistotyön asemaan liittyvä epätietoisuus on näkynyt organisoinnissa. Ennen vuotta 2012 kieliaineistotyö tapahtui pääosin tietohuolto-osastoon kuuluvassa arkistoyksikössä. Vuoden 2012 alussa lakiuudistuksen myötä organisaatio uudistettiin niin, että kieliaineistotyö sulautettiin muiden osastojen (kielenhuolto, sanakirjat, ruotsi) ja yksiköiden (hallinto, tietohallinto, kehittäminen) tehtäviin kuitenkin niin, että johtajan suorassa alaisuudessa työskenteli koordinoivassa roolissa aineistoasiantuntija. Vuoden 2014 alussa aineistotyön organisointia muutettiin uudelleen, ja perustettiin aineistoyksikkö. Vuonna 2016 kirjasto siirrettiin aineistoyksikön alaisuuteen, jolloin syntyi nykyinen kirjasto- ja aineistoyksikkö.

8 PÄÄTELMÄT

Kotuksen kieliaineistojen ja niiden parissa tehtävän työn tarkastelu osoittaa, että organisaatiolla on kirjattuja tietovarantojen avaamiseen ja digitalisaatioon liittyviä tavoitteita, jotka ovat linjassa toimintaympäristön odotusten kanssa. On myös määritelty toimenpiteitä, jotka tukevat tavoitteiden saavuttamista. Lisäksi on meneillään Nimiarkiston digitalisointihanke, joka toimii monessa suhteessa oppina ja suunnannäyttäjänä. Niin ikään henkilökunta on kyselyn perusteella motivoitunut.

Strategisia tavoitteita on kuitenkin edelleen saavuttamatta ja toimintatavat eivät kaikin osin tue varsinkaan projektimaisesti toteutuvaa digitaalisten aineistojen kehittämistä. Kokonaisarkkitehtuurimenetelmän hyödyntäminen voisi tuoda selkeyttä sekä strategisten tavoitteiden hallitsemiseen että toimintatapojen kehittämiseen.

Uudistunut JHS 179:n kokonaisarkkitehtuurimalli sisältää myös kyvykkyydet: ”Organisaatioiden kyvykkyyden nähdään koostuvan kolmesta osakokonaisuudesta: toimintamallit ja prosessit, henkilöstö ja osaaminen sekä tiedot ja järjestelmät. Usein näihin kolmeen kyvykkyyden osatekijään lisätään vielä kaksi myöhemmin tarkasteluun tullutta tekijää eli verkostoituminen ja uudistumiskyky.” Kotuksen kieliaineistojen näkökulmasta keskeisiä kyvykkyyksiä ovat seuraavat:

- digitaalisuuteen ja keskittymiseen perustuva toimintamalli (kaiken keskeisen digitalisointi, liepeellisen siirtäminen muualle)
- selkeät ja tehokkaat prosessit suurten kokonaisuuksien hallitsemiseksi
- digitalisoimisen hyödyt ja mahdollisuudet taitava henkilökunta
- selkeä, asiakaslähtöinen järjestelmä- ja palvelukokonaisuus
- vuorovaikutus ja yhteistyö kansallisen tason toimijoiden kanssa

Nykyinen kokoelma- ja palvelupolitiikka on osin vanhentunut, joten se tulisi päivittää.

Kotuksen projektitoiminta on kehittynyt kymmenen viime vuoden aikana, ja alkaa monin osin jo olla lähellä sitä, mitä projektitoiminnalla tätä nykyä tarkoitetaan. Projektikäsikirja tuo toivon mukaan edelleen parannusta tilanteeseen, ja auttaa juurruttamaan hyvää toimintakulttuuria. Yksittäiset onnistuneetkaan projektit eivät kuitenkaan yksin riitä koko organisaation näkökulmasta, sillä projektitoiminnan ja erityisesti kehittämissuunnitelman

voimavarat pitäisi suunnata strategian ja organisaation kehittämisen kannalta oikeisiin kohteisiin ja varmistaa projektityöllä valmisteltujen muutosten jalkauttaminen organisaation toimintaan. Tämä edellyttää organisaatiolta ja sen johtamiselta määrätietoisuutta ja muutosvalmiutta. Toiminnan kehittämisen kannalta esimerkiksi prosessien tunnistaminen ja kuvaaminen olisivat tärkeitä, nämäkin aiheet ovat olleet tunnistettujen kehittämistarpeiden listoilla jo vuosia ilman, että asiat ovat edenneet.

Kotuksen kannalta näyttää tarkoituksenmukaiselta, että mahdollisimman suuri osa aineistoista keskitetään Fin-Clarín-konsortion (ks. <http://www.kielipankki.fi>) tarjoamaan tieteen tietotekniikan keskuksen CSC:n ylläpitämään yhteiseen ympäristöön, joka koostuu kolmesta osasta:

- Korp-järjestelmä (<http://korp.csc.fi>) tekstikorpuksille
- LAT-järjestelmä (<http://lat.csc.fi>) av-korpuksille
- wiki-alusta sanastoaineistoille (<http://sanat.csc.fi>), ja tämän pohjalta ratkaisu digitaaliselle Nimiarkistolle
- Meta-Share (<http://metashare.csc.fi/>) metatiedoille

Lisäksi opetus- ja kulttuuriministeriön Avoin tiede ja tutkimus (ATT, ks. <http://www.avointiede.fi>) -hanke tarjoaa tiedon avaamiseen liittyviä ratkaisuja, joista tähän mennessä on otettu tai ollaan ottamassa käyttöön seuraavat:

- AVAA, avointen aineistojen julkaisualusta (<http://avaa.tdata.fi/>), nyt julkaistu paikkatietoaineistoja (rajapintapalvelut Geoserverin kautta)
- IDA, tutkimusaineistojen säilytyspalvelu (<http://avointiede.fi/ida>)
- Etsin, tutkimusdatan hakupalvelu (<https://etsin.avointiede.fi/>)

Myös opetus- ja kulttuuriministeriön hallinnonalalla toimiva Kansallinen digitaalinen kirjasto tarjoaa yhteisiä palveluja.

- Finna-hakupalvelu metatiedoille (<https://www.finna.fi/>)
- Pitkäaikaissäilytys digitaalisten aineistojen säilyttämiseen (PAS, <http://www.kdk.fi/index.php/fi/pitkaaikaissailytys>)

Edellä kuvattujen yhteisten ratkaisujen lisäksi ydintoimintojen eli sanakirjatyön ja kieltenhuollon tarpeisiin tarvitaan omia räätälöityjä ratkaisuja, koska näillä alueilla Kotus on ainoa toimija. Nykyisellään on käytössä sähköinen ympäristö laajojen sanakirjojen (mm.

Suomen murteiden sanakirja ja *Vanhan kirjasuomen sanakirja*) toimittamiseen ja verkkojulkaisemiseen. Kielenhuollon verkkopalvelut on puolestaan integroitu osaksi organisaation perusverkkosivustoa.

Jatkossa Kotus voisi keskittyä tietojärjestelmiin ja palveluihin, joita varten ei ole löydettävissä yhteisiä ratkaisuja:

- tieteellisten sanakirjojen verkkojulkaisu ympäristö (olemassa)
- (tausta)järjestelmä kuvailevalle ja hallinnolliselle metatiedoille (olemassa)

Alla oleva kuvio 4 esittää tiivistetysti nykyiset digitaalisten kieliaineistojen palvelukanavat. Uusissa järjestelmissä tulisi kiinnittää erityistä huomiota luvanvaraisten aineistojen käyttöön saattamiseen (mm. HAKA-autentikointi, tietoturva), joukkoistamisen ja talkootyön mahdollisuuteen sekä tietojen avaamiseen rajapintojen kautta. Uudet digitointi- ja muut aineistohankkeet tulee sovittaa määriteltyihin tietojärjestelmä- ja palveluympäristöihin ja välttää aineistokohtaisia erillisratkaisuja.

Kotus		KDK		ATT (CSC)			Fin-Clarin		
taustajärjestelmät (aineistojen metatiedot) ja rajapinnat	sähköiset sanakirjat	Finna-hakuliittymä	PAS-pitkäaikais-säilytys	IDA-säilytyspalvelu	Etsin-hakupalvelu	AVAA-julkaisu alusta	LAT-avokorpus alusta	Korpustekstikorpus alusta	Sanastosanasto alusta
Kotus/VALTORI		CSC:n tarjoamat tekniset yhteiset ja räätälöidyt ratkaisut (HAKA ym.)							

KUVIO 4. Kotuksen digitaalisten kieliaineistojen palvelukanavat ja niiden tekninen toteutus (alarivi).

LÄHTEET

Arkistolaki (831/1994). Finlex. <http://www.finlex.fi/fi/laki/ajantasa/1994/19940831>

Arkistowiki. <http://wiki.narc.fi/arkistowiki/index.php/Etusivu>

Avoim data – innovoinnin, kasvun ja läpinäkyvän hallinnon moottori. Euroopan komissio, 2011. <http://eur-lex.europa.eu/legal-content/FI/TXT/?uri=CELEX:52011DC0882>

Avoim tiede ja tutkimus. Käsitteet. Opetus- ja kulttuuriministeriö. Avoim tiede ja tutkimus -hanke. <http://avointiede.fi/kasikirja> > Käsitteet [käytetty 1.10.2014]

Borg, Sami & Kuula, Arja 2007: *Julkisrahoitteisen tutkimusdatan avoin saatavuus ja elinkaari. Valmisteluraportti OECD:n datasuosituksen toimeenpanomahdollisuuksista Suomessa*. http://www.fsd.uta.fi/fi/julkaisut/julkaisusarja/FSDjs06_OECD.pdf

Choo, Chun Wei 2001: *Information Management for the Intelligent Organization – The Art of Scanning the Environment*. 3rd ed. American Society for Information Science and Technology.

Creative Commons. <http://creativecommons.fi/> [käytetty 1.10.2014]

Ehdotus Kotuksen palautejärjestelmän toteuttamiseksi 13.6.2012. Palautejärjestelmäryhmä Kirsti Aapala, Salli Kankaanpää, Antti Leino, Elisa Stenvall. Helsinki: Kotimaisien kielten tutkimuskeskus.

EU:n tietosuoja-asetus (GPDR). <http://eur-lex.europa.eu/legal-content/FI/TXT/PDF/?uri=OJ:L:2016:119:FULL&from=FI>

Gustafsson, Sari & Marniemi, Janne 2012: *Julkisen johtamisen työkirja. Kuinka kasvaa menestyväksi organisaatioksi*. Helsinki: Tietosanoma.

Henkilötietolaki (523/1999). Finlex. <http://www.finlex.fi/fi/laki/ajantasa/1999/19990523>

INSPIRE-direktiivi (2007/2/EY). Euroopan unioni. <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002>

JHS 171 *ICT-palvelujen kehittäminen: Kehittämiskohteiden tunnistaminen*. <http://www.jhs-suositukset.fi/suomi/jhs171>

JHS 172 *ICT-palvelujen kehittäminen: Esiselvitys*. <http://www.jhs-suositukset.fi/suomi/jhs172>

JHS 179 *ICT-palvelujen kehittäminen: Kokonaisarkkitehtuurin kehittäminen*. <http://www.jhs-suositukset.fi/suomi/jhs179>

JHS 179 päivitys *ICT-palvelujen kehittäminen: Kokonaisarkkitehtuurin kehittäminen*. Päivitysluonnos 20.4.2015. <http://www.jhs-suositukset.fi/web/guest/projects/jhs-179-update/discussion-forum>

JHS 189 *Avoimen tietoaaineiston käyttöluupa*. <http://www.jhs-suositukset.fi/suomi/jhs189>

Julkisen hallinnon paikkatiedon viitearkkitehtuuri. <http://www.paikkatietoikuna.fi/web/fi/viitearkkitehtuuri>

Järvinen, Pekka T. & Rantala, Jukka & Ruotsalainen, Petri 2014: *Johda suoritusta*. Helsinki: Talentum.

Kansallisen digitaalisen kirjaston kokonaisarkkitehtuuri v. 2.0. <http://www.kdk.fi/images/tiedostot/KDK-kokonaisarkkitehtuuri2.pdf>

Kansallisen digitaalisen kirjaston standardisalkku. 27.10.2014. http://www.kdk.fi/images/tiedostot/KDK_standardisalkku27_10_2014.pdf

Kataisen hallitusohjelma 2011 = *Pääministeri Jyrki Kataisen hallituksen ohjelma 22.6.2011*. <http://valtioneuvosto.fi/documents/10184/147449/Kataisen+hallituksen+ohjelma/81f1c20f-e353-47a8-8b8f-52ead83e5f1a>

Kokoelma- ja palvelupolitiikka: kieliaineistot ja kirjasto. Helsinki: Kotimaisten kielten keskus, 2012.

Kotimaisten kielten keskuksen intranetsivut (jr = johtoryhmän pöytäkirjat, hankeohje 2007, hankelomakepohja 2007, hankelista 2008, täytetyt hankelomakkeet).

Kotimaisten kielten keskuksen strategia 2016–2020. Helsinki: Kotimaisten kielten keskus. https://www.kotus.fi/files/4947/Kotuksen_strategia_2016_2020.pdf

Kotimaisten kielten keskuksen suunnittelun ja seurannan asiakirjat (toimintakertomukset, tulossopimukset, toiminta- ja taloussuunnitelmat). http://www.kotus.fi/kotus/organisaatio/toiminnasta_raportointi ja http://www.minedu.fi/OPM/Tiede/ohjaus_ja_rahointus/tulosohjaus/?lang=fi

Kotimaisten kielten tutkimuskeskuksen tehtävät. Opetusministeriön työryhmämuistioita ja selvityksiä 2010:1. Helsinki: Opetusministeriö, 2010. <http://urn.fi/URN:ISBN:978-952-485-829-8>

Kotimaisten kielten keskuksen verkkosivut. <http://www.kotus.fi>

Kotimaisten kielten tutkimuskeskuksen palautejärjestelmä. Ehdotus 1.11.2011. Työryhmä Salli Kankaanpää, Raija Miikkulainen ja Elisa Stenvall. Helsinki: Kotimaisten kielten tutkimuskeskus

Laki Kotimaisten kielten keskuksesta (1403/2011). Finlex. <http://www.finlex.fi/fi/laki/alkup/2011/20111403>

Marvia, Jaana 2005: Kirjoitetun kokoelmapolitiikan laatimisesta. OTA-kirjasto E 4. Espoo: Teknillinen korkeakoulu. <http://lib.tkk.fi/Julkaisut/E4.pdf> [käytetty 1.10.2014]

Ojasalo, Katri & Moilanen, Teemu & Ritalahti, Jarmo 2014: *Kehittämistyön menetelmät. Uudenlaista osaamista liiketoimintaan*. 3., uud. p. Helsinki: Sanoma Pro Oy.

Opetus- ja kulttuuriministeriön ja Kotimaisten kielten keskuksen välinen tarkistettu tulossopimus vuosille 2017–2020 ja voimavarat vuodelle 2017.

Perustietovarantojen viitearkkitehtuuri. <https://www.avoindata.fi/data/fi/dataset/perusta-perustietovarantojen-viitearkkitehtuuri-1-0>

Poikola, Antti & Kola, Petri & Hintikka, Kari A. 2010: *Julkinen data – johdatus tietovarantojen avaamiseen*. Helsinki: Liikenne- ja viestintäministeriö. <http://www.julkinen-data.fi/>

PSI-direktiivi (2003/98/EY) ja direktiivi PSI-direktiivin muuttamisesta (2013/37/EU). Euroopan unioni. <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32003L0098> | <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:02003L0098-20130717>

Salokannel, Marjut 2012: *Tutkimuksen perusteet: Tutkimusdatan ja muun primäärin tutkimusaineiston saatavuuteen liittyvä oikeudellinen sääntely*. Edilex. https://helda.helsinki.fi/bitstream/handle/10138/39575/Salokannel_Tutkimuksen_perusteet.pdf?sequence=2

Salokannel, Marjut 2013: *Tekijänoikeus ja tutkimuksen raaka-aineet: Uuden teknologian haasteet tekijänoikeudellisesti suojatun materiaalin tutkimuskäytölle*. Helsinki: Koneen säätiö. www.koneensaatio.fi/index.php/download_file/view/752/244/

Sinkkonen, Irmeli & Kuoppala, Hannu & Parkkinen, Jarmo & Vastamäki, Raino 2006: *Käytettävyyden psykologia*. 3., uudistettu p. Helsinki: Edita.

Sipilän hallitusohjelma 2015 = *Valtioneuvoston tiedonanto eduskunnalle 29.5.2015 nimitetyn pääministeri Juha Sipilän hallituksen ohjelmasta*. http://valtioneuvosto.fi/documents/10184/1433371/Tiedonanto_Sipil%C3%A4_29052015_final.pdf/6de03651-4770-492a-907f-89452141d0d5

Stubbin hallitusohjelma 2014 = *Valtioneuvoston tiedonanto eduskunnalle 24.6.2014 nimitetyn pääministeri Alexander Stubbin hallituksen ohjelmasta*. <http://valtioneuvosto.fi/documents/10184/145135/Stubbin+hallituksen+ohjelma/fafd39bc-307c-4cde-8b4c-b97724cf24ef>

Suutari, Toni 2006a: Kotimaisten kielten tutkimuskeskuksen sähköiset aineistot 22.3.2016. Kotimaisten kielten tutkimuskeskus.

Suutari, Toni 2006b: Sähköiset hetket – Kotuksen aineistoja verkkoon. – *Kielikello* 4/2006, s. 6–9. Helsinki: Kotimaisten kielten tutkimuskeskus.

Suutari, Toni 2015: Kotimaisten kielten keskuksen digitaalisten kieliaineistojen kohdearkkitehtuurin nykytila. Kotimaisten kielten keskus.

Suutari, Toni 2016: Kohti avoimia ja osallistavia aineistopalveluja. – *Kielikello* 2/2016, s. 4–6. Helsinki: Kotimaisten kielten keskus.

Tekijänoikeuslaki (404/1961). Finlex. <http://www.finlex.fi/fi/laki/ajantasa/1961/19610404>

Tieteen termipankki. <http://tieteentermipankki.fi>

Tietohallintolaki = *Laki julkisen hallinnon tietohallinnon ohjauksesta* (634/2011). Finlex. <https://www.finlex.fi/fi/laki/ajantasa/2011/20110634>

TKS = *Tietojenkäsittelyn kehittämissuunnitelma* 1987–1991. Helsinki: Kotimaisten kielten tutkimuskeskus, 1986.

Toiminta- ja taloussuunnitelma 2014–2017. Opetus- ja kulttuuriministeriön julkaisuja 2013:1. Helsinki, 2013. <http://www.minedu.fi/OPM/Julkaisut/2013/tts.html?lang=fi>

Toiminta- ja taloussuunnitelma 2017–2020. Kotimaisten kielten keskus.

Toiminta- ja taloussuunnitelma 2018–2021. Kotimaisten kielten keskus.

Tuomo, Tuomi 1975: Sanakirjasäätiö. – Eila Hämäläinen & Kaisu Juusela (toim.): *Castrenianum. Suomen ja sen sukukielten tutkimuskeskus*, s. 9–26. (Castrenianumin toimitteita 10.) Helsinki.

Valtioneuvoston asetus Kotimaisten kielten keskuksesta (1566/2011). Finlex. <http://www.finlex.fi/fi/laki/alkup/2011/20111566>

Valtioneuvoston asetus Kotimaisten kielten keskuksesta (1566/2011). Finlex. <http://www.finlex.fi/fi/laki/alkup/2011/20111566>

Valtioneuvoston periaatepäätös julkisen sektorin digitaalisten tietoaineistojen saataavuuden parantamisesta ja uudelleenkäytön edistämisestä 3.3.2011. http://www.lvm.fi/c/document_library/get_file?folderId=1551281&name=DLFE-11828.pdf

Wikipedia. <http://fi.wikipedia.org> [käytetty 1.10.2014]

Virtanen, Petri 2006: *Houkutteleva työyhteisö*. Helsinki: Edita.

VMBaro-järjestelmän verkkosivut. <http://www.baro.vm.fi/>

LIITTEET

Liite 1. Termit ja lyhenteet

Seuraavassa on määritelty Kotuksen sähköisten kieliaineistojen kannalta keskeisimmät käsitteet sekä hankkeet ja palvelukokonaisuudet. Muuten terminologiaan voi perehtyä esimerkiksi Kansallisen digitaalisen kirjaston kokonaisarkkitehtuuriin monikielinen sanaston (<http://www.kdk.fi/fi/kokonaisarkkitehtuuri/sanasto>), ja Avoimen tieteen käsikirjan sanaston (<http://avointiede.fi/kasikirjan-sanasto>) avulla.

AVAA

Avoimien tieteiden ja tutkimuksen -hankkeen tarjoama avointen aineistojen julkaisualusta (<http://avaa.tdata.fi/>).

Avoimien tieteiden ja tutkimuksen (ATT)

Avoimien tieteiden ja tutkimuksen -hanke on opetus- ja kulttuuriministeriön vuonna 2014 käynnistämä hanke, joka edistää tiedon avointa saatavuutta (<http://avointiede.fi>). Hanke tarjoaa organisaatioille ja tutkijoille tietoa, koulutusta ja palveluja. Yhteisten palvelujen avulla digitaalisia tutkimusaineistoja voidaan hallinnoida, avata ja jakaa muiden hyödynnettäväksi sekä pitkäaikaissäilyttää luotettavasti.

avoimien data/tieto

Avointa dataa ja tietoa käytetään limittäin, mutta joissakin yhteyksissä *datalla* voidaan tarkoittaa raakadataa ja *tiedolla* raakadataa jalostuneempaa tietoa, vrt. englannin *open data* ja *open knowledge*. Avoimella tarkoitetaan avointa käytettävyyttä, ei pelkästään tietojen julkaisuutta (<http://www.hri.fi/fi/mita-on-avoimien-data/>). Seuraavat määritelmät löytyvät opetus- ja kulttuuriministeriön Avoin tiede ja tutkimus -hankkeen verkkosivuilta (<http://avointiede.fi/kasikirjan-sanasto>):

- ”Avoimella tiedolla viitataan digitaalisessa muodossa oleviin sisältöihin ja dataan, joita kuka tahansa voi vapaasti ja maksutta käyttää, muokata ja uudelleen jakaa. Avoimen tiedon määritelmän mukaisesti aineiston tulee olla kokonaisuudessaan saatavilla käyttökelpoisessa ja muokattavassa muodossa Internetin kautta ja sen tulee olla lisensoitu niin, että sen käyttöä, muokkausta ja uudelleenjakelua ei rajoiteta.”

- ”Avoimella datalla tarkoitetaan tutkimukselle, julkishallinnolle, organisaatioille, yrityksille tai yksityishenkilöille kertynyttä jalostamatonta informaatiota, joka on avattu ulkopuolisillekin vapaasti ja maksutta hyödynnettäväksi.”

CAF = Common Assessment Framework

CAF on EU-jäsenmaiden yhteistyönä suunnittelema julkisen sektorin organisaatioiden laadunarviointimalli (www.valtiokonttori.fi/caf).

CC = Creative Commons

CC lisenssejä käytetään maailmanlaajuisesti avoimia aineistoja julkaistaessa (<https://creativecommons.fi>). Suomessa julkishallinnon avoimissa tietoaaineistoissa suositellaan käytettäväksi Creative Commons Nimeä 4.0 -käyttöluvaa (<http://www.jhs-suositukset.fi/suomi/jhs189>).

digitalisointi

Digitalisoinnilla ei tarkoiteta pelkästään esimerkiksi mekaanista dokumenttien skannausta kuvatiedostoiksi, vaan siihen kuuluvat aineistojen elinkaarimallin eri osa-alueet (ks. kuvaa 2 s. 13). Usein digitalisoinnin yhteydessä joudutaan parantamaan metatietojen tasoa ja hallintaa sekä arvioimaan aineiston käyttöön liittyviä juridisia, eettisiä ja sopimuksellisia asioita sinänsä ja ratkaisemaan, miten näitä tietoja säilytetään ja hyödynnetään tietojärjestelmissä (esim. käyttäjänhallintaan integroituina). Aineiston digitoiminen, pitkäaikaissäilytys, käyttöön saattaminen ja jatkokäytön edistäminen ovat samoin jokainen kokonaisarkkitehtuurin kannalta keskeisiä osa-alueita. Osaan aineistoista liittyy tekijänoikeuteen ja henkilösuojaan liittyviä käytön rajoituksia, joten myös käyttäjänhallintaa ja tietoturvaa joudutaan pohtimaan monessa yhteydessä.

Dublin Core

Dublin Core on digitaalisten aineistojen kuvailun metatietoformaatti (<http://dublin-core.org/>).

Etsin

Avoin tiede ja tutkimus -hankkeen tarjoama tutkimusdatan hakupalvelu (<https://etsin.avointiede.fi/>).

fenistiikka

Fennistiikalla tarkoitetaan suomen kielen tieteellistä tutkimusta (Tieteen termipankki).

Fin-Clarín ks. Kielipankki

Finna

Finna (www.finna.fi) on Kansallisen digitaalisen kirjaston (KDK) asiakasliittymä, joka on otettu tuotantokäyttöön lokakuussa 2013. Hakupalvelu tarjoaa keskitetyn pääsyn Suomen arkistojen, kirjastojen ja museoiden aineistojen metatietoihin, ja osaan aineistoista pääsee Finnasta suoraan.

IDA

Avoin tiede ja tutkimus -hankkeen tarjoama tutkimusaineistojen säilytyspalvelu (<http://avointiede.fi/ida>).

Kansallinen digitaalinen kirjasto (KDK)

Opetus- ja kulttuuriministeriön Kansallinen digitaalinen kirjasto -hanke (KDK, www.kdk.fi) edistää kirjastojen, arkistojen ja museoiden digitaalisten aineistojen saatavuutta ja käytettävyyttä. Keskeisimmät palvelut ovat asiakasliittymä Finna (www.finna.fi) ja pitkäaikaissäilytysjärjestelmä (PAS). Hanke edistää mukana olevien organisaatioiden yhteistyötä ja yhteisten ratkaisujen käyttöönottoa. Finna- ja PAS-palvelut jatkavat pysyvinä kansallisina palveluina hankkeen päätyttyä vuoden 2017 lopussa.

Kielipankki

Kielipankki tarkoittaa Fin-Clarín-konsortion tarjoamaa palvelukokonaisuutta (www.kielipankki.fi), jonka kautta digitaalisia kieliaineistoja tarjotaan keskitetysti käytettäväksi. Fin-Clarín toimii Helsingin yliopiston johdolla ja siihen kuuluvat suomalaiset yliopistot, Kotimaisten kielten keskus sekä Tieteen tietotekniikan keskus CSC, joka vastaa palvelun teknisestä ylläpidosta. Fin-Clarín on osa eurooppalaista Clarín Eric-konsortiota, jonka tavoitteena on kehittää digitaalisten ihmistieteiden yhteistä tutkimusinfrastruktuuria.

kokoelma- ja palvelupolitiikka

Kokoelmapolitiikat ovat tuttuja erityisesti kirjastoista ja museoista, mutta niitä käytetään myös arkistoissa. Politiikassa käsitellään kokoelmien muodostamista, ylläpitoa, arvioin-

tia ja kehittämistä. Kirjastojen näkökulmasta asiaa on käsitelty mm. Marvia (2005, käsitteiden määrittelyä erit. s. 6–8). Palvelupolitiikka linjaa organisaation tarjoamia palveluja, niiden laatua, kattavuutta ja toimintatapaa. Yrityksissä palveluja voidaan kuvata myös esimerkiksi palvelustrategian, -kuvauksen ja -lupauksen avulla. Palveluihin liittyviä käsitteitä on määritelty mm. Tekesin Palveluliiketoiminnan sanastossa.

kokonaisarkkitehtuuri

Julkisen hallinnon JHS 179 -suositus ICT-palvelujen kehittäminen: Kokonaisarkkitehtuurin kehittäminen (s. 9) määrittelee kokonaisarkkitehtuurin seuraavasti:

- ”Kokonaisarkkitehtuuri on toiminnan, prosessien ja palvelujen, tietojen, tietojärjestelmien ja niiden tuottamien palvelujen muodostaman kokonaisuuden rakenne. Kokonaisvaltainen lähestymistapa organisaation toiminnan ja sen rakenteiden hallinnoimiseksi ja kehittämiseksi.”

Kehittämistyössä keskeistä ei ole kaiken kokonaisarkkitehtuuriin kuuluvan yksityiskohdainen kuvaaminen vaan kokonaisuuden ja sen kehittämistarpeiden hahmottaminen. Tarkastelussa ei pitäydytä organisaatiotasolla, koska yhä enemmän toimitaan yhteistyössä eri hankkeissa ja konsortioissa ja koska kokonaisarkkitehtuurityötä tehdään valtionhallinnossa laajemmin sektoreittain, esimerkiksi muistiorganisaatioiden (kirjastot, arkistot, museot) kokonaisarkkitehtuurityö on alkamassa.

Korp

Kielipankin tarjoama tekstikorpusten julkaisemiseen ja käyttöön tarkoitettu järjestelmä (<http://korp.csc.fi>).

korpus

Kielitieteessä korpus on ”suhteellisen laaja, järjestelmällisesti kerätty kokoelma näytteitä luonnollisesta kirjoitetusta tai puhutusta kielestä” (Tieteen termipankki).

LAT

Kielipankin tarjoama av-korpusten julkaisemiseen ja käyttöön tarkoitettu järjestelmä (<http://lat.csc.fi>).

lisensointi

Lisensointi liittyy keskeisesti tiedon avaamiseen, koska valtaosa sisällöistä on tekijänoikeuden ja sen lähioikeuksien suojaamaa. Avointen sisältöjen lisensoinnilla sallitaan teoksen vapaa käyttö, kopiointi ja edelleenlevittäminen, usein myös kaupallisissa tarkoituksissa. Tiedon lisensoinnissa tärkeimpiä ovat CC eli Creative Commons -lisenssit (<http://creativecommons.fi/>).

Meta-Share

Eurooppalainen kieliaineistojen metatiedoille tarkoitettu järjestelmä (<http://metashare.csc.fi/>), josta Suomessa vastaa Kielipankki.

muistiorganisaatio

Muistiorganisaatioita ovat kirjastot, arkistot ja museot.

Nimiarkisto

Nimiarkisto on yksi Kotimaisten kielten keskuksen arkistoista. Arkiston kokoelmissa on kattavat tiedot Suomen paikannimistä ja sen lisäksi muita nimikokoelmia (mm. henkilönimiä ja vanhoista asiakirjoista poimittuja historiallisia nimiä). Digitaalinen Nimiarkisto (www.nimiarkisto.fi) avautuu vuoden 2017 lopussa, kun suomenkielisten paikannimien kokoelmat tulevat verkon kautta käytettäviksi.

pitkäaikaissäilytys (PAS)

Digitaalisen tiedon pitkäaikaissäilytyksellä (<http://www.kdk.fi/fi/pitkaaikaissailytys>) tarkoitetaan informaation luotettavaa säilyttämistä vuosikymmenten ja jopa vuosisatojen ajan, vaikka laitteet, ohjelmistot ja tiedostomuodot vaihtuvat. Kansallinen digitaalinen kirjasto -hankkeessa on luotu mukana oleville organisaatioille yhteinen, kansallinen PAS-palvelu.

provenienssiperiaate

Provenienssiperiaatteen ”mukaan asiakirjatiedot kuuluvat sen yhteisön tai henkilön arkistoon, jonka toiminnan tuloksena ne ovat syntyneet tai joiden haltuun ne ovat asian käsittelyn yhteydessä päätyneet”. Käytännössä periaate tarkoittaa sitä, että esimerkiksi keskus- tai päätearkistoon siirrettäessä eri arkistonmuodostajien aineistoja ei sekoiteta keskenään. Periaatteeseen kuuluu myös se, että aineistot pyritään säilyttämään alkuperäisessä järjestyksessä. (Arkistowiki.)

Sanat

Kielipankin tarjoama wiki-alusta sanastoaineistoille (<http://sanat.csc.fi>).

talkoistaminen/joukkoistaminen

Joukkoistamisella ja talkoistamisella tarkoitetaan toimintamallia, jossa ongelmanratkaisu tai muu tekeminen on hajautettu suurelle joukolle ihmisiä (<http://fi.wikipedia.org/wiki/Joukkouttaminen>). Joukkoistamalla on kerätty jo ennen digitaalista aikaa tutkimus- ja kulttuuriperintöaineistoja; digitaalisena aikana menetelmä tarjoaa uudenlaisia mahdollisuuksia.

Liite 2. Kieliaineistojen kokonaisarkkitehtuuri. Luettelo ohjaavasta tiedosta

Lainsäädäntö

- *Henkilötietolaki*
- *Laki Kotimaisten kielten keskuksesta*
- *Laki tekijänoikeudesta*
- *Laki viranomaisten toiminnan julkisuudesta*

EU:n asiakirjat

- *INSPIRE-direktiivi (2007/2/EY)*
- *PSI-direktiivi (2003/98/EY) ja direktiivi PSI-direktiivin muuttamisesta (2013/37/EU)*
- *EU:n tietosuoja-asetus (GDPR)*

Normit

- *Arkistolaitoksen suositus digitoinnin laatuksiteereiksi 2008*
- *JHS 171 ICT-palvelujen kehittäminen: Kehittämiskohteiden tunnistaminen*
- *JHS 172 ICT-palvelujen kehittäminen: Esiselvitys*
- *JHS 179 ICT-palvelujen kehittäminen: Kokonaisarkkitehtuurin kehittäminen*
- *JHS 189 Avoimen tietoaaineiston käyttö lupa*
- *Kansallisen digitaalisen kirjaston standardisalkku*

Kansalliset asiakirjat

- *Valtioneuvoston periaatepäätös julkisen sektorin digitaalisten tietoaaineistojen saatavuuden parantamisesta ja uudelleenkäytön edistämisestä [2011]*

Kotuksen omat asiakirjat

- *Kotimaisten kielten keskuksen strategia 2012–2015 [2012]*
- *Kokoelma- ja palvelupolitiikka: kieliaineistot ja kirjasto [2012]*
- *Toiminta- ja taloussuunnitelmat [vuosittain tehtävä]*
- *Tulossopimukset OKM:n ja Kotuksen välillä [vuosittain tehtävä]*

Viitearkkitehtuurit

- *Arkaluonteisen tutkimuskäytön viitearkkitehtuuri (luonnos 8.4.2015)*

- *Arkistosektorin aineistohallinnan viitearkkitehtuuri*
- *Julkisen hallinnon paikkatiedon viitearkkitehtuuri*
- *Kansallisen digitaalisen kirjaston kokonaisarkkitehtuuri*
- *Perustietovarantojen viitearkkitehtuuri*