

OKM:n hallinnonalan julkisten tietovarantojen avaaminen - selvitys suunnitelmista toimialoittain

Korkeakoulu- ja tiedepolitiikka

Korkeakoulu- ja tiedepolitiikan (KTPO) alueella keskeiset rahoitusta kaudella 2014-2017 vaativat hankkeet kustannuksineen on esitetty alla olevassa taulukossa. Hankkeet tuottavat tietovarantojen avaamiseen ja hyödyntämiseen liittyvää osaamista ja ratkaisuja, joilla on laajaa poikkihallinnollista vaikuttavuutta ja huomattava potentiaali yksityisen ja julkisen T&K&I-työn tehostamiseen. Tarkempi kuvaus hankkeista löytyy liitteistä.

	Tyyppi*	2014	2015	2016	2017	Yhteensä
Selvitys tietovarantojen hyödyntämisen problematiikasta yksityisen ja julkisen T&K&I-työn tehostamisen näkökulmasta (mom. 29.40.20)	1,2,3,4	300	300			600
AVOIN - avoimen tiedon instrumentit (a. aineistojen avaaminen, b. osaamisen tukeminen, c. prosessien kehittäminen, d. pilotit) (mom. 29.40.20)	1,2,3,4	360 0	360 0	240 0	2400	12000
Kotimaisten kielten keskuksen Nimiarkiston paikannimikokoelmat (mom. 29.40.03)	2	350	350	200	200	1100
Avataan vaiheittain Kotimaisten kielten keskuksen rajoituksettomat kieliaineistot (mom. 29.40.03)	2	240	240	240	240	960
Korkeakoulujen metatietopääoman avaaminen (mom. 29.40.20)	1,2,3,4	800	600	300	200	1900
Arkistolaitoksen ns. projektitietokannat (Suomen sotasurmat 1914-1922 ym.) (mom. 29.40.02)	2	55				55
Yhteensä		534 5	509 0	314 0	3040	16415

* Tietovarannon tyyppien vaihtoehdot:

- 1) *jo avattu (saatavissa koneluettavassa muodossa maksutta)*
- 2) *julkinen, avattavissa oleva, johon ei liity mitään käyttörajoituksia tai lainsäädännön esteitä avaamiselle*
- 3) *tietovaranto, joka mahdollisesti voitaisiin ja voisi olla hyödyllistä avata, mutta jonka avaamiseen liittyvät kysymykset vaativat erillistä selvittelyä*
- 4) *salassa pidettävä, joka voidaan jättää tämän selvityksen ulkopuolelle*

Hankkeissa syntyvää osaamista ja kehitettyä tekniikkaa voidaan hyödyntää muilla hallinnonaloilla ja yritystoiminnassa. Kehittämisen lähtökohtana on yhteentoimivien laajasti käytettävissä olevien ratkaisujen kytkeminen erityisosaamisiin. Hankkeissa luodaan muun muassa yhtenäinen paikkatiedon käsittelyn tapa ja menetelmät avattujen tietoaaineistojen kytkemiseen karttatietopalveluihin.

Arkistolaitoksella on merkittäviä ja laajoja tietovarantoja, joihin liittyviin Arkistolaitoksen ehdotuksiin tarvitaan selvitystä elinkaarikustannuksista ja yhteentoimivuuden ratkaisusta, ja tätä työtä tehdään AVOIN-hankkeen yhteydessä. Tarvittavien selvitysten jälkeen voidaan arvioida esitysten toteuttamisen vaatimia resursseja ja kytkentää muuhun kehitystoimintaan. Varastokirjaston ja Suomen Akatemian osalta ei esitetä lisämäärärahan tarvetta.

Yllä olevassa taulukossa mainittuihin kehityshankkeisiin kytkeytyy erillinen ehdotus pitkäaikaissäilytysjärjestelmästä digitaalisille kulttuuriperintöaineistoille ja tutkimuksen tietoaaineistoille. Kansallinen digitaalinen kirjasto -hankkeessa (KDK) ja Tutkimuksen tietoaaineistot -hankkeessa (TTA) valmistettava pitkäaikaissäilytysjärjestelmä (KDK-PAS ja TTA-PAS) vähentää digitaalisten kulttuuriperintöaineistojen ja tutkimuksen tietoaaineistojen hallinnan, jakelun ja säilyttämisen päällekkäisiä toimia ja pidemmällä aikavälillä kustannusten nousua. Vuosien 2014-2017 aikana toteutettavan hankkeen lisämäärärahan tarve on 34,7 miljoonaa euroa. Hanke sisältyy tutkimus- ja innovaatiopolitiikan toimintaohjelman (TINTO) luonnokseen.

Esitetyt toimenpiteet tehostavat sekä yksityisen että julkisen sektorin palveluinnovaatioita ja edistävät tiedon yhteiskäyttöä ja avoimuutta. Hankkeet tehostavat muilla hallinnonaloilla käynnistettyjä kehittämistoimia (mm. LVM:n FORGE-pilvipalvelulaboratorio ja VM:n Avoimen tiedon ohjelma) tuomalla parhaan tietoaaineistoihin liittyvän tutkimusosaamisen ja ratkaisut tarjolle kaikille hallinnonaloille. Yritystoiminnan kannalta toimenpiteillä on poikkeuksellisen suurta pitkän aikavälin potentiaalia.

Kulttuuri-, liikunta- ja nuorisopolitiikka

Kulttuurisektorille kuuluvien valtion virastojen ja laitosten (Kansallinen audiovisuaalinen arkisto, Museovirasto, Näkövammaisten kirjasto, Suomenlinnan hoitokunta, Valtion taidemuseo, Taiteen keskustoimikunta) tietovarantojen avaamista edistävät toimenpiteet on esitetty kustannuksineen alla olevassa taulukossa.

Toimenpiteistä koituvia hyötyjä ovat ajantasaisen viranomaistiedon (mm. maankäytön suunnittelu ja kaavoitus) ja tietovarantojen alueellisesti tasavaroisen saatavuuden parantuminen, uusien innovatiivisten käyttömuotojen mahdollistuminen, tietovarantojen avaamisen mahdollistama päällekkäisen tiedontuotannon karsiminen, kulttuuristen sisältöjen hyödyntäminen digitaalisissa tuotteissa ja palveluissa sekä uuden tiedon tuottaminen (tutkimuskäyttö, innovaatio- ja kehitystoiminta) ja opetuskäyttö.

Kulttuurisektorin virastojen ja laitosten tietovarantojen avaamista ei voida toteuttaa kehyksen puitteissa. Seuraavassa on esitetty arvio aineistojen avaamisesta aiheutuvasta lisämäärärahan tarpeesta tietovarannoittain ja palveluittain. Esitettyjä lisäyksiä ei ole kohdennettu OKM:n TTS-esityksessä momenteille. Toteuttaminen edellyttää lisämäärärahan osoittamista joko keskitetysti tai kyseisten virastojen ja CSC:n toimintamenoihin.

- Kulttuurisektorin virastojen ja laitosten tietovarantojen avaamisesta koituvat, lisämäärärahan tarpeen arvioinnissa huomioonotetut kustannukset koostuvat seuraavista toimenpiteistä:
- kulttuurin metatieto- ja tietovarantojen yhteisen latauspalvelun toteuttaminen
- digitaalisten metatieto- ja tietovarantojen saattaminen koneluettavaan muotoon ja keskeisten aineistojen digitoiminen
- palvelurajapintojen suunnittelu ja toteuttaminen
- palvelukapasiteetin riittävyyden turvaaminen virastoissa ja laitoksissa (tietojen käytön volyymin kasvu, käyttömuotojen monipuolistuminen)

Lisäksi kulttuurisektorin laitoksille ja virastoille koituu tietovarantojen avaamisesta seuraavia kustannuksia, joita ei ole otettu lisämäärärahan tarpeen arvioinnissa huomioon:

- mahdollisista maksutulojen menetykset
- tietovarantojen jatkojalostamisen edellyttämien tekijänoikeuksien hankkiminen

Kulttuurisektorin toimijoiden tietovarantojen avaamiseen kytkeytyy saumattomasti Kansallinen digitaalinen kirjasto -hankekokonaisuus (KDK), jolla tuodaan kirjastojen, arkistojen ja museoiden digitaaliset aineistot ja palvelut helposti saataville (KDK-asiakasliittymä) ja varmistetaan kulttuuriperintöaineistojen luotettava pitkäaikaissäilytys (KDK-PAS). KDK- ja TTA-hankkeista on toimitettu TTS-prosessiin erillinen muistio (LIITE 2).

OKM / Kulttuurisektorin virastojen ja laitosten tietovarantojen avaaminen aineistotyypeittäin 2013 - 2017							
Tietovarannot	aineis- to- tyyp- pi* (1,2,3)	Avaamisvu- osi / avattu	2013	2014	2015	2016	2017
Yhteinen latauspalvelu kulttuurisektorin virastojen ja laitosten tietovarantojen avaamiseksi koneluettavassa muodossa.	1,2,3	2013/2014	200 000	100 000			
Kulttuuriympäristön tietovarannot (Museovirasto) sekä Suomenlinnan hoitokunnan tietovarannot	1	Avattu osin (latauspalvelu toiminnassa). Tietovarantojen kertaluonteinen avaaminen 2013. Sisältöjen laajentaminen 2014 - 2016	100 000	250 000	250 000	250 000	
Esinekokoelmien metatiedot ja osa digitaalisista objekteista (Museovirasto) sekä taideteosten metatiedot ja osa digitaalisista objekteista (Valtion taidemuseo)	1,2,3	Avattu osin (Finna ja Europeana). Sisältöjen laajentaminen 2013 - 2017		450 000	450 000	350 000	350 000
Kirjastojen metatietovarannot / Näkövammaisten kirjasto, Museoviraston kirjasto ja Valtion taidemuseon	1,2	Museoviraston kirjaston metatietovaranto avataan 2013. Muut avattavissa 2014 - 2016.		50 000	50 000	50 000	

kirjasto							
Kansatieteen ja historian kuvakokoelmat (Museovirasto) sekä kuvataiteen keskusarkiston valokuvakokoelmat, kuvataiteen ja taidehistorian arkistotietokannat ja AV-tietokanta (Valtion taidemuseo)	1,2,3	Avattu osin (Finna ja Europeana). Sisältöjen laajentaminen ja avaaminen vaiheittain 2014 - 2017.		250 000	250 000	250 000	250 000
Elokuvien filmografiset tiedot (Kansallinen audiovisuaalinen keskus KAVA)	2	Avattavissa 2013	20 000				
Apuraha- ja määräraharekisteri sekä hankerekisteri (Taiteen keskustoimikunta, 1.1.2013 alkaen Taiteen edistämiskeskus)	2,3	Avattavissa 2013 - 2014					
Yhteensä / vuosi			320 000	1 100 000	1 000 000	900 000	600 000

***Tyyppien selitteet:** 1) jo avattu (saatavissa koneluettavassa muodossa maksutta); 2) julkinen, avattavissa oleva, johon ei liity mitään käyttörajoituksia tai lainsäädännön esteitä avaamiselle; 3) tietovaranto, joka mahdollisesti voitaisiin ja voisi olla hyödyllistä avata, mutta jonka avaamiseen liittyvät kysymykset (esim. tekijänoikeus) vaativat erillistä selvittelyä ja oikeuksien hankkimista

Koulutuspolitiikka

Koulutuspolitiikan osaston toimialan suunnitelmat on esitelty liitteessä 10.

Liitteet

- Liite 1: KTPO:n virastojen ja laitosten suunnitelmat sekä muu käytetty taustamateriaali
- Liite 2: Selvitys tietovarantojen hyödyntämisen problematiikasta yksityisen ja julkisen T&K&I-työn tehostamisen näkökulmasta
- Liite 3: AVOIN - avoimen tiedon instrumentit
- Liite 4: Kotimaisten kielten keskuksen (Kotus) Nimiarkiston paikannimikokoelmat
- Liite 5: Kotimaisten kielten keskuksen (Kotus) rajoituksettomat kieliaineistot avataan vaiheittain 2013-2017
- Liite 6: Korkeakoulujen metatietopääoman avaaminen
- Liite 7: Arkistolaitoksen ns. projektitietokannat (Suomen sotasurmat 1914-1922 ym.)
- Liite 8: KUPO:n virastojen ja laitosten tietovarantojen avaaminen
- Liite 9: Digitaalisten kulttuuriperintöaineistojen ja tutkimuksen tietoaineistojen pitkäaikaissäilytysjärjestelmän toteuttaminen

- Liite 10: Koulutuspolitiikan osaston toimialan suunnitelmat

Liite 1: KTPO:n virastojen ja laitosten suunnitelmat sekä muu käytetty taustamateriaali

Digitaalisten kulttuuriperintöaineistojen ja tutkimuksen tietoaineistojen hallinnan, jakelun ja säilyttämisen tietoinfrastruktuuria kehitetään Kansallinen digitaalinen kirjasto -hankkeessa (KDK) ja Tutkimuksen tietoaineistot -hankkeessa (TTA) siten, että palvelut ja tiedot ovat saatavilla ja käytettävissä helposti ja turvallisesti.

KTPO:n osalta tietovarantojen avaamisen suunnitelma perustuu TTA- ja KDK-hankkeissa tehtyyn suunnittelutyöhön sekä tietohallintolaissa määritetyn OKM:n kohdealueen koulutus, tiede ja kulttuuri piirissä järjestettyihin seminaareihin ja tapaamisiin, joissa tietovarantojen avaamista on käsitelty. Lisäksi laitoksilta ja virastoilta pyydettiin marraskuussa 2012 tietoa niiden tietovarannoista OKM:n suunnitelman laatimista varten.

OKM:n alaisten virastojen ja laitosten tietovarannoille tyypillistä, että ne ovat vielä fyysisessä muodossa. Fyysisten kokoelmien täysimittainen digitointi on monessa suhteessa problemaattista. Pitkäaikaisten kustannusten arviointi on haastavaa ja osaaminen on rajallista.

Digitointihankkeissa on välttämätöntä toimia OKM:n kohdealueen yhteentoimivuutta edistävien hankkeiden puitteissa (esimerkiksi TTA- ja KDK-hankkeet). Tietoarkkitehtuurin rakentaminen on ensisijaisen tärkeää, jotta kokonaisuutta voidaan hallita ja kustannustaso pysyy hallussa. Käytännössä tarvitaan myös merkittäviä panostuksia tietojärjestelmäkehitykseen ja tallennuskapasiteettiin.

Kokonaisarkkitehtuurinäkökulma on tulevaisuudessa kehittämissä hankkeissa avainasemassa, sillä tällä hetkellä virastoilla on paljon toisistaan riippumatta kehitettyjä ratkaisuja ja rajapintoja. Yhteentoimivuuden kehittäminen vaatii aktiivista yhteistyötä toimialan organisaatioiden kesken.

Viranomaisten käsittelemällä tiedolla on nykyisessä toimintaympäristössä useita muitakin muotoja kuin asiakirja tai rekisteri, mikä nostaa esille kysymyksen siitä, ovatko ”paperiaikakaudella” kehitetyt prosessit ja toimintamallit ylipäänsä relevantteja digitaalisessa toimintaympäristössä. Tähän liittyen on tarkoituksenmukaista muun muassa selvittää arkistolain uudistamistarvetta.

Kohdealueella tarvitaan merkittävässä määrin yhteistä suunnittelu- ja kehitystyötä jotta voidaan taata kehitettävien ratkaisujen yhteentoimivuus. Lisäksi on tarpeen selvittää digitoitavaksi ehdotettujen aineistojen elinkaarikustannuksia ennen kuin niiden osalta voidaan päättää toimenpiteistä.

Arkistolaitoksen tekemien ehdotusten osalta tarvitaan selvitystä elinkaarikustannuksista ja yhteentoimivuuden ratkaisuksista, ja tätä työtä tehdään AVOIN-hankkeen yhteydessä. Kun tarvittavat selvitykset on tehty, voidaan arvioida esitysten toteuttamisen vaatimia resursseja ja kytkentää muuhun

kehitystoimintaan. Varastokirjaston ja Suomen Akatemian osalta ei esitetä lisämäärärahan tarvetta.

Arkistolaitos:

- Astia-palvelun metatiedot (tyyppi 1 ja 2)
- Arkistolaitoksessa säilytettävät ja sinne siirrettävät paperiaineistot (tyyppi 3)
- Digitaaliarkiston aineisto (tyyppi 2)
- Suomen sotasurmat 1914-1922, Suomen sodissa menehtyneet 1939-1945, Suomi, sotavangit ja ihmislouvatukset, Diplomatarium Fennicum ja Heraldica (Ns. projektitietokannat) (tyyppi 2)
- Sähköisen aineiston vastaanotto- ja palvelujärjestelmä VAPA (tyyppi 3)
- Aarre-arkistorekisteri (tyyppi 3)

Suomen Akatemia:

- Tutkimusrahoitushakemukset (tyyppi 4)
- Rahoituspäätökset (tyyppi 1 ja tyyppi 3)
- Tutkimusrahoitusta koskevat loppuraportit (tyyppi 4)
- Suomen Akatemian toteuttamien tutkimusjärjestelmätason arviointien loppuraportit ja aineistot (tyyppi 1 ja tyyppi 3)

Kotimaisten kielten keskus (Kotus):

- Kotimaisen kielten keskuksen rajoituksettomat kieliaineistot (tyyppi 2)
- Kotimaisten kielten keskuksen Nimiarkiston paikannimikokoelmat (tyyppi 2)

Varastokirjasto:

- Vaari, Bibliografinen tietokanta kirjaston aineistosta (tyyppi 1)

Tietovarannon tyyppien vaihtoehdot:

- 5) *jo avattu (saatavissa koneluettavassa muodossa maksutta)*
- 6) *julkinen, avattavissa oleva, johon ei liity mitään käyttörajoituksia tai lainsäädännön esteitä avaamiselle*
- 7) *tietovaranto, joka mahdollisesti voitaisiin ja voisi olla hyödyllistä avata, mutta jonka avaamiseen liittyvät kysymykset vaativat erillistä selvittelyä*
- 8) *salassa pidettävä, joka voidaan jättää tämän selvityksen ulkopuolelle*

Liite 2: Selvitys tietovarantojen hyödyntämisen problematiikasta yksityisen ja julkisen T&K&I-työn tehostamisen näkökulmasta

Yksityisen ja julkisen T&K&I-työn tehostaminen edellyttää selvitystä tietovarantojen hyödyntämisen problematiikasta, jotta panostukset voidaan kohdistaa vaikuttavuuden kannalta keskeisimpiin tietovarantoihin ja niiden ympärille rakentuviin ekosysteemeihin.

Vuosina 2014-2015 tehtävässä selvityksessä paikallistetaan osaamisen, menetelmien ja tietoteknisten ratkaisujen pullonkauloja pyrkien lisäämään toimenpiteiden vaikuttavuutta sekä yksityisen että julkisen sektorin puolella ja tunnistamaan olennaisia kehittämiskohteita niin osaamisen kuin tietoaineistojen yhteentoimivuuden ja hyödyntämisessä tarvittavien tietoteknisten ratkaisujen alueella.

Digitaalisten kulttuuriperintöaineistojen ja tutkimuksen tietoaineistojen hallinnan, jakelun ja säilyttämisen tietoinfrastruktuuria kehitetään Kansallinen digitaalinen kirjasto -hankkeessa (KDK) ja Tutkimuksen tietoaineistot -hankkeessa (TTA) siten, että palvelut ja tiedot ovat saatavilla ja käytettävissä helposti ja turvallisesti.

Vuosille 2011-2013 käynnistetty TTA-hanke panostaa tutkimuksen tietoaineistojen hyödyntämiseen liittyvän tahtotilan vahvistamiseen ja kansallisen tietopolitiikan luomiseen sekä tutkimuksen tietoinfrastruktuurin rakentamiseen. Hankkeen aikana mm. rakennetaan federoitu tutkimusaineistojen tallennuspalveluratkaisu, tuetaan metatiedon tuottamista ja yhdenmukaistetaan tutkimusaineistojen tuottamiseen ja ylläpitämiseen liittyviä prosesseja sekä selvitetään Kansallisen digitaalisen kirjaston kanssa yhteinen pitkäaikaistallennusratkaisu.

Hankkeen taustalla on vuosina 2009-2010 käynnissä ollut Tutkimuksen tietoaineistot -selvityshanke sekä sen pohjalta vuonna 2011 valmistunut selvitys *Tieto käyttöön. Tiekartta tutkimuksen sähköisten tietoaineistojen hyödyntämiseksi*.

Vuosina 2014-2015 tehtävässä selvityksessä hyödynnetään TTA- ja KDK-hankkeissa luotua poikkihallinnollista osaamista tietoaineistojen avaamisen tematiikasta ja luodaan parhaaseen käytettävissä olevaan tietämykseen perustuva arvio keskeisistä pullonkauloista niin osaamisessa, menetelmissä kuin tietoteknisissä ratkaisuissakin.

Liite 3: AVOIN - avoimen tiedon instrumentit

Tutkimuksen tuottamaa osaamista tarvitaan, jotta avatut tietoaineistot tulevat hyötykäyttöön ja niiden innovaatiopotentiaalia voidaan maksimaalisesti hyödyntää.

Korkeakoulujen rooli Suomen tutkimus- ja innovaatiojärjestelmässä on keskeinen sekä uuden tiedon tuottamisessa että saatavilla olevan tiedon jalostamisessa innovaatioiksi niin yhteiskunnan kuin yritystoiminnan hyödyksi. OKM:n suunnitelmassa luodaan edellytyksiä sille, että korkeakoulujen tuottamaa osaamista voidaan parhaalla mahdollisella tavalla käyttää tietovarantojen hyödyntämisessä. Samalla tuetaan yritysten kykyä kehittää liiketoimintaansa.

Yksityisen ja julkisen sektorin pitkäjänteiseen T&K&I-työhön tulee panostaa, mikä edellyttää hyvin toimivaa infrastruktuuria, erityisesti tietoinfrastruktuuria. Jotta tietovarantojen avaamisesta on hyötyä pitemmällä tähtäimellä, tulee tutkimuksen tietoinfrastruktuuria kehittää kokonaisuutena huomioiden yhteentoimivuuden vaatimukset poikki hallinnonalojen.

Tietohallintolain mukaisesti OKM koordinoi yhteentoimivuutta koulutuksen, tieteen ja kulttuurin kohdealueella, mikä tarkoittaa että OKM:n suunnitelmassa tulee huolehtia laajemmin kuin oman hallinnonalan osalta tietovarantojen avaamiseen liittyvistä yhteentoimivuuden kysymyksistä. OKM:n vastuulla olevaan kohdealueeseen sisältyy tutkimus, jonka rooli on keskeinen julkishallinnon tietovarantojen avaamisessa ja hyödyntämisessä yhteiskunnan käyttöön, mukaan lukien yksityisen sektorin tarpeet.

Tiedon liikkuvuutta tutkimusinfrastruktuurien, korkeakoulujen ja muiden sidostoimijoiden välillä voidaan tukea kehittämällä yhteistyörakenteita, tietomäärittelyjä ja tietojärjestelmiä. Tiedon yhteismitallisuuden ja yhteentoimivuuden kehittäminen koskee koulutusta ja tutkimusta sekä korkeakoulujen hallintoa.

Digitaalisten kulttuuriperintöaineistojen ja tutkimuksen tietoaineistojen hallinnan, jakelun ja säilyttämisen tietoinfrastruktuuria kehitetään Kansallinen digitaalinen kirjasto -hankkeessa (KDK) ja Tutkimuksen tietoaineistot -hankkeessa (TTA) siten, että palvelut ja tiedot ovat saatavilla ja käytettävissä helposti ja turvallisesti.

OKM:n kehittämishankkeissa (esimerkiksi TTA ja KDK) tuetaan tietovarantojen avaamista ja yhteentoimivuutta. Näiden hankkeiden kautta on mahdollista laajentaa myös kansallista palveluarkkitehtuuria, koska tietyt avainteknologiat (kuten pitkäaikaissäilytys) vaativat erityisosaamista, jota ei laajalti löydy. Aineistojen tuominen tarjolle hyödyntämiskelpoisella tavalla (palvelualusta, tiedostomuodot, rajapinnat jne.) vaatii jatkossa kannusteita ja tukea. Erityisenä teemana nousee esille paikkatieto, joka on lähes kaikissa OKM:n virastojen ja laitosten tietovarannoissa keskeistä ja samalla hallinnonaloja yhdistävää.

Tiedon avaamisessa myös tietoturvakysymykset ovat tärkeitä. Tietoturvasojen aukoton määrittäminen tarvitaan niin tutkimus- kuin viranomaistoiminnassa aineistojen sujuvan hyödyntämisen varmistamiseksi

Riittävän yhteismitallisuuden ja yhteentoimivuuden saavuttaminen edellyttää pitkäjänteistä yhteistyön tiivistämistä ja muun muassa lainsäädännön uudistamista. Toimenpiteet ovat korkeakoulujen sekä opetus- ja kulttuuriministeriön yhteisiä ja yhdessä koordinoitavia nopean ja tarkan toimeenpanon mahdollistamiseksi.

Neljä keskeistä kehittämisaluetta Avoimen tiedon instrumentit (AVOIN) -hankekokonaisuudessa

Seuraavien kehittämisalueiden avulla tehostetaan avattujen tietoaineistojen laajaa hyödynnettävyyttä ja mahdollistetaan tietoaineistojen tehokas yhdistely (ns. digitaalinen tutkimusprosessi):

- a) Avataan tutkimusaineistoja hyödynnettäviksi
- b) Tuetaan tiedon tuottamiseen ja hyödyntämiseen liittyvää osaamista
- c) Kehitetään tiedon tuottamisen prosesseja
- d) Avoimen tiedon hyödyntämisen pilotit.

a) Avataan tutkimusaineistoja hyödynnettäviksi

Tiedon avaamisen ja hyödyntämisen kannalta on keskeistä, että koulutuksen, tieteen ja kulttuurin kohdealueella on käytössä hyvin ylläpidetty, yhteentoimiva ja tehokas tietoinfrastrukturi, joka voidaan määritellä seuraavasti: tietojen sijaintiin, rakenteeseen, organisointiin, hakemistoihin ja luetteloihin, omistajuuteen, saatavuuteen, varmistukseen, tietoturvaan ja tietovarastoihin liittyvät keskeiset resurssit ja kyvykkyydet sekä elinkaaren hallinta.

Tietojärjestelmien ja yhteistyörakenteiden kehittämiseksi tarvitaan yhteinen kieli tietojen kuvaamiseksi, eli yhteismitallinen käsitteistö ja ratkaisuja, jotka mahdollistavat tiedon liikkuvuuden organisaatioiden välillä. Käsitteistön tulee noudattaa soveltuvia kansainvälisiä standardeja.

Sanastoja tulee hyödyntää kuvailun yhtenäistämisen ja metatiedon automaattisten kielikonversioiden mahdollistamiseksi. Yhteismitallisen tiedon liikuttamiseksi tarvitaan yhteentoimivia rajapintaratkaisuja, tietovarantoja ja tietojärjestelmiä.

Koko julkishallinnon kattavien toimialariippumattomien metatietopalveluiden lisäksi tarvitaan toimialariippuvia palveluita. Seuraavassa on hahmoteltu OKM:n toimialan tarvitsemia ratkaisuja, joita ei muualla synny:

- avointen aineistojen julkaisun alusta, joka mahdollistaa helpon tavan saattaa tutkimuksen tietoaineistot avoimesti hyödynnettäväksi
- tutkimuksen tarpeita palvelevan avoimen rajapintapalvelun kehitystyö
- rajapintaratkaisun rakentaminen mm. julkishallinnon perustietovarantojen hyödyntämiseen (esim. VIA-liittymä)
- auktoriteettitiedon palvelu (ISNI, ORCID, toimijarekisteri)
- aineistojen viittauskäytäntöjä ja meritoitumista tukeva palvelu (esim. DataCite-liittymä, DOI).

Kehittämisaikavälillä rakennetaan avoimen datan julkaisun alusta, joka mahdollistaa semanttisen tietosisällön tuotannon sekä sisällön yhdistämisen toisiin tietosisältöihin yhteentoimivasti mm. Linked data -lähestymistavalla. Tällöin voidaan toteuttaa datan julkaiseminen helposti käyttöön otettavina verkkopalveluina (mashup, web service). Kohdeainestoina ovat sekä a) tutkimuksen tuottamat tietoaaineistot että b) tutkimuksen tarvitsemat tietoaaineistot.

Ratkaisulla tuetaan korkeakoulujen mahdollisuuksia tuottaa aineistojensa ja niiden metatiedot uudelleenkäyttöön sopivassa formaatissa sekä kehitetään sovelluksia ja palveluita aineistoja hyödyntämään. Tukeminen voidaan tehdä avoimen julkaisun alustan yhteydessä, tukitoiminnon muodossa. Lisäksi perustetaan tiedon saatavuuden klinikka tukemaan, ohjeistamaan ja neuvomaan tietovarantojen avaamisen kysymyksissä.

Alustaan liitetään seuraavat työkalut:

- metatiedon tuotannon automatisointi
- tiedon laadun ja luotettavuuden tunnistaminen
- metadatajoukkojen yhdistäminen
- siltaukseen liittyvät palvelut rajapintoihin.

Toteutuksessa hyödynnetään tuloksia ja toimitaan yhteistyössä muun muassa Aallon LDF-hankkeen kanssa. Kehitystyö tehdään Aallossa, ja palvelu siirtyy tuotantoon CSC:lle osaksi TTA-hankkeen palveluita. Avoimen datan julkaisualusta on toteutus TTA-hankkeen tietoinfrastruktuurityöryhmän tiedon tuotannon prosessiin liittyvästä julkaisupalvelusta.

Tutkimustiedon avaamisen alusta kattaa seuraavat osiot:

- 1) avoimien aineistojen tuottamisen työkalut
- 2) julkishallinnon avattujen aineistojen haku ja yhdistely tutkimuskäyttöön
- 3) sensitiivisten aineistojen turvallisen tutkimuskäytön kokonaisuus
- 4) kansalaistieteen näkökulma.

Alustaan rakennetaan yleiskäyttöinen käyttöoikeuksien hakupalvelu. Toimialariippuvassa sovelluksessa käyttäjä tunnistetaan HAKA:lla. Aineistojen oikeudenhaltijat voivat palvelun avulla myöntää aineistoon sähköisesti käyttöoikeuden. Lisäksi huomioidaan ei-HAKAan tms. luottamusverkostoon kuuluvat henkilöt, esim. aineistojen oikeudenhaltijat. Palvelussa voi olla erilaisia käyttäjärooleja. Aineiston metatiedon yhteyteen talletetaan tieto mm. siitä, missä sijaitsee kyseistä aineistoa koskeva alkuperäinen tallennussopimuspaperi (oikeudenhaltijan kanssa tehty).

Toteutuksessa pyritään myös yhtenäistämään ja ajantasaistamaan tietoaaineistojen tuotantoympäristöjä, jotta arvokkaat tietoaaineistot turvataan ja niiden käyttökelpoisuus pysyy vaatimusten tasalla.

Sensitiivisten aineistojen turvallista etäkäyttöä edistetään luomalla alusta, johon aineistontuottajat voivat tietoturvallisesti antaa rajapinnan kautta käyttöoikeuksia aineistoihinsa, sekä virtuaalisia etätyöpöytiä tutkimuskysymysten ratkaisuun hallinnoiva työskentelykokonaisuus. Virtuaalietätyöpöytiä tarvitaan niin rekisteritutkimuksen, kielentutkimuksen (FIN-CLARIN) kuin biolääketieteen käyttöön (BBMRI).

b) Tuetaan tiedon tuottamiseen ja hyödyntämiseen liittyvää osaamista

Osaamisen tukemiselle tarvitaan eri tasoisia toimenpiteitä. Kansainvälisten huippuosaajien toiminta vaatii jatkossa osaamista liittyen aineistojen analysointiin, tiedon siirtoon, tiedon louhintaan, tiedon pakkaukseen, hyödyntämiseen, yhdistämiseen, harmonisointiin, etc. Tutkimusasetelmat muuttuvat, ja uusi tarve tulee olemaan esimerkiksi aineisto-arkeologia, jossa selvitetään aineiston historiaa. Kun koko tutkimusprosessi digitalisoituu, tarvitaan mm. tätä aineisto-arkeologiaa tutkimuksen verifiointia ja validointia varten.

Voimavaroja tulee suunnata erityisosaamisen rakentamiseen digitaalisten palveluiden, prosessien ja avainteknologioiden alueille. Tällaista erityisosaamista kaivataan optimoituihin tiedonsiirtoratkaisuihin, tietoturvakysymyksiin, luottamusverkostojen ja lupaprosessien hallintaan, mobiliteettiin jne. Voimakas avainteknologia on ns. big data -kenttä, joka parhaimmillaan yhdistää suurten datamäärien tallennuksen (pilvitekniologia) ja tuottamisen (sensorit) innovatiiviseen sovelluksiin ja analyysityökaluihin.

Kokonaisuuksien hallintaan tarvitaan erityisosaamista, jonka kasvattaminen on tulevaisuuden kannalta kriittistä. Osaaminen yhdistää tietotekniset taidot kommunikointi- ja organisointikykyyn sekä erityisalojen (esimerkiksi ympäristöosaaminen tai metalliteollisuus) substanssin ymmärrykseen.

c) Kehitetään tiedon tuottamisen prosesseja

Tietoinfrastruktuuria hyödyntävien prosessien tukeminen tehostaa tutkimus- ja innovaatiotoimintaa. Tällaisia toiminnan tehokkuuden kannalta keskeisiä prosesseja ovat metatiedon tuottaminen (sanastojen hyödyntäminen, käsittemallien tuottaminen), tietovirtojen ohjaus, rajapintatoteutuksien hyödyntäminen, prosessikuvauksien tuottaminen ja järjestelmäintegraatiot.

Vahva tiedon tuottamisen prosesseja ja sitä kautta tiedon saatavuutta ja laadukkuutta edistävä toimenpide on temaattisten keskittyminen luominen ja tuleminen. Temaattisella keskittymällä tarkoitetaan tiettyyn tutkimukselliseen asiakokonaisuuteen keskittyvää yhteistoimintaa, kuten tutkimusinfrastruktuurin.

Temaattisten keskittymien perustana ovat substanssiosaajien tietovarastoihin tuottamat tietoaineistot. Tietovarastojen elinkaaren hallinnan kannalta tietoaineistojen kattava kuvaaminen on välttämättömyys. Tutkimuksen tuottamien tietoaineistojen suhteen on epäsuhta tietoaineistojen tuottamiseen käytettävän suhteellisen suuren resurssin suhde aineistojen kuvailuun ja hallintaan käytettyyn pieneen resurssin. Temaattiset keskittymät voivat ottaa

vastuuta asiakokonaisuuksien tietovarastoista, tiedon ymmärrettävyydestä ja hallinnasta.

d) Avoimen tiedon hyödyntämisen pilotit

Avoimen tiedon hyödyntämisen pilotit tuottavat käytännön kokemusta tietovarantojen avaamisen toteuttamisesta. Samalla saadaan näyttöä onnistumisista.

Seuraavien toimenpiteiden kautta voidaan pilotoida mm. joukkoistamista aineistojen avoimen julkaisemisen kautta:

- Best practises -tyyppinen menetelmäpilotti, jossa ”rikastetaan” hyödyntämiskelpoiseksi valittu datajoukko Tim Berners-Leen 5-tähden mallin mukaisesti. Pilotissa kehitettävän lähestymistavan hyötyä demonstroidaan sovelluksilla. Tämä kehittää tutkimuksellista vetovoimaa tietoaaineistojen avaamiselle. Vuonna 2013 laaditaan kansallinen tiekartta tutkimusinfrastruktuureista, ja tämä tiekartta määrittää kansallisen infrastruktuuripolitiikan painopisteet ja tutkimusongelmien ympärille rakentuvat temaattiset keskittymät, joiden toimintaa tuetaan menetelmäpilotin avulla. Pilotin tulokset ja ratkaisut voidaan avata suoraan näiden syntyvien temaattisten keskittymien käyttöön tiedon avaamisen avuksi. Pilottiaineistoina tietoaaineistojen yhdistelyssä tutkimustarkoituksiin voivat olla seuraavat aihealueet:
 - o hiukkasfysiikka: CERNin CMS-kokeesta syntyneet tutkimusaineistot.
 - o luonnontieteelliset havaintotietokannat biodiversiteetin hallintaan
 - o Ilmatieteen laitoksen (IL) säädata.
- Paikkatietoaineistot tutkimusaineistojen erityiskohteena, mutta myös yrityspuolen tarpeiden tukemisessa: esim. OpenStreetMap -rajapinta avoimen paikkatietorajapinnan testaamiseksi.
- Avattujen tutkimuksen tietoaaineistojen käyttäminen koulujen ja lukioiden opetuksessa ja kansalaistieteessä (esimerkiksi hiukkasfysiikan CMS-tutkimusdata).

Liite 4: Kotimaisten kielten keskuksen (Kotus)

Nimiarkiston paikannimikokoelmat

Kotuksen aineistot ovat kansallisen tason tutkimusinfrastruktuuri (ks. OKM:n julkaisuja 1/2009). Nimiarkisto kokoaa ja ylläpitää suomen- ja saamenkielisiä nimiaineistoja tutkimuksen ja nimistönhuollon tarpeisiin. Nimiarkistossa on noin 2,7 miljoonan nimilipun verran paikannimitietoa. Lisäksi tietoa on sidoksina ja käsikirjoituksina. Nimilipussa on yleiskielistetty nimen hakuasu, nimen murremuoto ja taivutus, sijainti (pitäjä, kylä, talo ja karttasignumi), paikan laji, mahdollisesti selitys sijainnista sekä kaikki haastatteleamalla saatu nimeä tai paikkaa valaiseva muistitieto. Nimilipussa olevien sanallisten tietojen lisäksi nimet on paikannettu keruukarttoihin (n. 10 000 kpl) tehdyin merkinnöin.

Arkistossa on arvokasta tietoa paitsi nimistön- ja kielentutkijoille myös arkeologeille, historiantutkijoille, kansatieteilijöille ja luonnontieteilijöille. Myös sukututkijoille arkisto tarjoaa kiinnostavaa oheis- ja taustamateriaalia. Nimiin liittyvät tiedot voidaan avata osin rajoituksetta (mm. nimet, sijainnit) ja osin rajoitetusti (mm. henkilöjä koskeva muistitieto).

Kotuksen lakisäätteiset tehtävät eli suomen ja ruotsin kielten huolto, neuvonta ja sanakirjatyö sekä kielenhuoltoon ja sanakirjatyöhön liittyvä tutkimus edellyttävät, kuten lain perusteluissa todetaan, "kieliaineistojen ja -arkistojen ylläpitoa ja kehittämistä". (Laki Kotimaisten kielten kesuksesta 1566/2011.) Kotuksen kieliaineistoihin ei sovelleta arkistolakia, koska Kotus vastaa kieliaineistoista tehtäviensä perusteella erityisviranomaisena.

Kotuksen aineistot ovat maksutta käytössä verkossa ja arkistoissa. Maksua peritään ainoastaan tietopalvelusta, kopioimisesta, digitaalisen aineiston toimittamisesta aiheutuvista kuluista yms., esim. kun tiedostoja toimitetaan asiakkaan tilauksen perusteella cd/dvd-levyssä. Maksukäytännöt perustuvat OKM:n asetukseen. (Asetus Kotimaisten kielten suoritteiden maksuista 1293/2010 ja 1475/2011.)

Tietovaranto voidaan avata sähköisessä muodossa digitoinnin jälkeen. Digitointia on suunniteltu eri yhteyksissä, ja vähäisiä osia Nimiarkistosta on digitoitu kuntien ym. tilauksesta sekä omana työnä. Saamelaisnimistön osalta (Kotuksessa n. 13 000 nimilippua, 750 keruukarttaa ym.) on tehty yhteistyössä eri organisaatioiden kanssa 2010–2011 hankesuunnitelma, joka on Biologista monimuotoisuutta koskevan yleissopimuksen alkuperäiskansojen perinnetietoa käsittelevän artikla 8j:n kansallisen asiantuntijatyöryhmän loppuraportin (<http://www.ymparisto.fi/default.asp?contentid=386848&lan=FI>) liitteenä nro 7 (<http://www.ymparisto.fi/download.asp?contentid=127695&lan=fi>). Hanke ei ole saanut rahoitusta, mutta osoittaa paikannimiaineiston digitoimisen tärkeyttä.

Nimiarkiston digitoiminen kokonaisuutena olisi tehokas ja yhteiskunnallisesti vaikuttava tapa toimia. Näin tietovaranto olisi kaikkien tarvitsijoiden käytettävissä. Paikannimitieto palvelee laajasti kansalaisten, tutkijoiden ja julkishallinnon tarpeita. Paikannimitietoa on eri toimijoilla, joten digitointi

mahdollistaisi eri organisaatioiden aineistojen yhteiskäytön. Paikannimitiedot ovat tärkeä osa paikkatietojen kokonaisuutta, ja ne ovat mukana paikkatietoinfrastruktuurilaissa.

Nimiarkiston avaaminen sähköisessä muodossa edellyttää aineiston digitointia (lippujen skannaus kuvamuotoon, sidosten ja keruukarttojen digitointi mikrofilmeistä). Massadigitointityö voidaan teettää parhaiten ostopalveluna. Digitoitu aineisto edellyttää asiantuntijatyönä vähintään perusmetatietojen liittämistä (esim. nimi, kunta, paikan laji) ja koordinaattipaikannusta (digitoitujen keruukarttojen sitominen koordinaatistoon ja paikannusten yhdistäminen nimitietoihin). Olennaista on myös se, että näin syntyvän paikkatietotietokannan tiedot ovat yhdistettävissä muihin paikkatietoihin. Noudatetaan INSPIRE-yhteensopivia ratkaisuja.

Osa Nimiarkiston nimitiedoista on yhteistä MML:n Paikannimirekisterin kanssa, eli on kyse samojen paikkojen nimistä. Nämä nimet tunnistetaan ja niiden metatietoihin sisällytetään viittaus Paikannimirekisterin paikka-id:hen. Nimilippujen sisältö on suurelta osin käsin kirjoitettua, joten automaattisen tekstintunnistuksen käyttömahdollisuudet ovat varsin rajalliset. Teknisissä ratkaisuissa ja yhteisissä käytännöissä Kotus toimii osana OKM:n hallinnonalan Fin-Clarín-konsortiota.

Näin ollen tulee huolehtia myös Fin-Clarínin riittävästä resursoinnista, jotta se kykenee muun muassa tarjoamaan HY:ssä koordinoitua ja CSC:n kautta tarvittavan teknisen infrastruktuurin. Aineiston saaminen käyttöön edellyttää digitoinnin ja tiedon käsittelyn lisäksi käyttöliittymää ja siihen liittyviä muita teknisiä ratkaisuja. Hyvän pohjan Nimiarkiston digitoinnille luo vuonna 2013 Kotuksessa toteutuva avoin data ja paikkatieto -hanke. Myös aiemmissa Kotuksen Paikannimipankki- ja Paikannimikartasto-hankkeissa on kertynyt hyödyllistä kokemusta, vaikka näissä hankkeissa on käsitelty vain hyvin pientä osaa koko aineistosta.

Nimiarkiston digitoinnin kokonaiskustannuksiin vaikuttaa olennaisesti se, missä laajuudessa nimilippujen sisältöä muunnetaan sähköiseksi tekstiksi. Alustava kustannusarvio Nimiarkiston avaamiseksi digitaalisessa muodossa on 800 000 euroa. Kustannukset muodostuvat digitoinnin ostopalveluista, palkkakustannuksista ja digitaalisen aineiston pitkäaikaissäilytyksen järjestämisestä. On varauduttava myös laite- ja ohjelmistokuluihin sekä toimitilakustannuksiin.

Kotuksen kieliaineistot ovat nykyiselläänkin maksuttomassa käytössä, joten avaamisesta ei aiheudu tulonmenetyksiä. Avaamisesta koituvat hyödyt ovat merkittävät. Sekä yliopistollinen että kaupallinen kieliteknologinen ja muu tutkimus tarvitsevat koneluettavia avoimia kieliaineistoja. Kieliteknologia on avainasemassa, kun kehitetään sovelluksia ja palveluja, jotka turvaavat suomen kielen asemaa digitaalisella aikakaudella (ks. Suomen kieli digitaalisella aikakaudella, <http://www.meta-net.eu/whitepapers/e-book/finnish.pdf>). Tutkimuksen ja kehityksen lisäksi aineistot hyödyttävät suoraan ja välillisesti opiskelijoita, opettajia, julkishallintoa ja myös laajasti kansalaisia. Avoimuus luo

uusia mahdollisuuksia esimerkiksi oppimisympäristöjen, sähköisten oppimateriaalien ja oppimissovellusten laatimiseen. Kieliaineistot tarjoavat aineksia perusopetuksesta yliopistotasolle asti. Kansalaisten kannalta kiinnostavia voisivat olla paikkakuntien kieltä, historiaa ja kulttuuriperinnettä esille tuovat sovellukset, jotka voisivat yhdistellä eri organisaatioiden aineistoja. Avaaminen tukee osallistumista ja osallistamista, joten olemassa olevien aineistojen avaamisen myötä on luontevaa etsiä keinoja uuden täydentävän aineiston keräämiseen (joukkoistaminen, kansalaistiede), onhan suuri osa nykyisistäkin valtavista kieliaineistoista talkootyönä aikanaan kerättyjä.

Liite 5: Kotimaisten kielten keskuksen (Kotus) rajoituksettomat kieliaineistot avataan vaiheittain 2013-2017

Kotuksen aineistot ovat kansallisen tason tutkimusinfrastruktuuri (ks. OKM:n julkaisuja 1/2009). Kotuksen aineistopalvelu Kainossa (<http://kaino.kotus.fi>) olevat aineistot voidaan avata myös koneluettavassa muodossa vuosina 2014-2017. Tänä aikana verkossa voidaan julkaista myös toistaiseksi julkaisemattomia, vielä käsittelyä vaativia aineistoja ja joitakin uusia, syntyviä aineistoja.

Kaino sisältää vapaasti käytettäviä aineistoja, kuten suomen kielen tekstikorpuksia 1500-luvulta 1800-luvulle, sananparsia, sanaston taajuustietoja, sana- ja nimiluetteloja, paikannimikartaston sekä erilaisia muita tietokantoja, jotka sisältävät sekä kielenaineiksia että kokoelmien metatietoja. (Avaaminen ei koske verkkosanakirjoja.)

Vuonna 2013 Kotus avaa erillishankkeessa INSPIRE-yhteensopivasti paikkatietoaineistoja eli alun perin kartastoina ilmestynyttä tai aiemmin julkaisematonta kartoiksi tuotettua tietoa murteiden sanastollisista ja kieliopillisista piirteistä, murrealuejaoista sekä murreosanaston ja paikannimistön levikeistä (kielinä suomi, muita itämerensuomalaisia kieliä ja suomenruotsi).

Kotuksen lakisäätteiset tehtävät eli suomen ja ruotsin kielten huolto, neuvonta ja sanakirjatyö sekä kielenhuoltoon ja sanakirjatyöhön liittyvä tutkimus edellyttävät, kuten lain perusteluissa todetaan, "kieliaineistojen ja -arkistojen ylläpitoa ja kehittämistä". (Laki Kotimaisten kielten kesuksesta 1566/2011.) Kotuksen kieliaineistoihin ei sovelleta arkistolakia, koska Kotus vastaa kieliaineistoista tehtäviensä perusteella erityisviranomaisena.

Verkossa olevat aineistot ovat maksutta käytössä. Maksua peritään ainoastaan tietopalvelusta, digitaalisen aineiston toimittamisesta aiheutuvista kuluista yms., esim. kun tiedostoja toimitetaan asiakkaan tilauksen perusteella cd/dvd-levyssä. Maksukäytännöt perustuvat OKM:n asetukseen. (Asetus Kotimaisten kielten suoritteiden maksuista 1293/2010 ja 1475/2011.)

Tietovaranto avataan vaiheittain 2013-2017. Vuonna 2013 Kotus avaa paikkatietoaineistojaan avoin data ja paikkatieto -hankkeessaan OKM:n erillisrahoituksella ja tekstikorpuksia siirretään Fin-Clarín-konsortion ylläpitämälle Korp-alustalle, joka sisältää rajapinnan. Vuosina 2014-2017 voidaan avata käytössä jo olevat rajoituksettomat kieliaineistot kuten myös uusia aineistoja, jos avaamiseen saadaan tarvittavat lisäresurssit 3 htv/vuosi.

Avaaminen edellyttää Kotukselle vuosille 2014-2017 lisäresurssiksi 3 htv/vuosi kieliteknologiseen, tietotekniseen ja sisällölliseen asiantuntijuuteen, jotta aineistoja voidaan käsitellä mm. standardien ja yhteisten käytäntöjen mukaisiksi (yhteensopivuus), jotta voidaan ottaa käyttöön rajapinta- ja tiedostonlatauspalvelut ja jotta voidaan tarjota tukea käyttäjille ja hyödyntäjille, esimerkiksi tukea sovelluskehittäjiä ja eri organisaatioiden tietovarantoja

innovatiivisesti yhdistäviä toimijoita, jotta avaamisesta koituu mahdollisimman paljon lisäarvoa ja vaikuttavuutta.

Teknisissä ratkaisuissa ja yhteisissä käytännöissä Kotus toimii osana OKM:n hallinnonalan Fin-Clarín-konsortiota. Näin ollen tulee huolehtia myös Fin-Clarínin riittävästä resursoinnista, jotta se kykenee muun muassa tarjoamaan HY:ssä koordinoitua ja CSC:n kautta tarvittavan teknisen infrastruktuurin.

Kotuksen rajoituksettomat kieliaineistot ovat nykyiselläänkin maksuttomassa käytössä, joten avaamisesta ei aiheudu tulojenmenetyksiä. Avaamisesta koituvat hyödyt ovat merkittävät. Sekä yliopistollinen että kaupallinen kieliteknologinen ja muu tutkimus tarvitsevat koneluettaavia avoimia kieliaineistoja. Kieliteknologia on avainasemassa, kun kehitetään sovelluksia ja palveluja, jotka turvaavat suomen kielen asemaa digitaalisella aikakaudella (ks. Suomen kieli digitaalisella aikakaudella, <http://www.meta-net.eu/whitepapers/e-book/finnish.pdf>). Tutkimuksen ja kehityksen lisäksi aineistot hyödyttävät suoraan ja välillisesti opiskelijoita, opettajia, julkishallintoa ja myös laajasti kansalaisia.

Avoimuus luo uusia mahdollisuuksia esimerkiksi oppimisympäristöjen, sähköisten oppimateriaalien ja oppimissovellusten laatimiseen. Kieliaineistot tarjoavat aineksia perusopetuksesta yliopistotasolle asti. Kansalaisten kannalta kiinnostavia voisivat olla paikkakuntien kieltä, historiaa ja kulttuuriperinnettä esille tuovat sovellukset, jotka voisivat yhdistellä eri organisaatioiden aineistoja. Avaaminen tukee osallistumista ja osallistamista, joten olemassa olevien aineistojen avaamisen myötä on luontevaa etsiä keinoja uuden täydentävän aineiston keräämiseen (joukkoistaminen, kansalaistiede), onhan suuri osa nykyisistäkin valtavista kieliaineistoista talkootyönä aikanaan kerättyjä.

Rajoituksettomien kieliaineistojen osalta avaamisella ei ole olennaisia vaikutuksia lainsäädännön kehittämistarpeisiin. Joissakin tapauksissa avaamista voi kuitenkin rajoittaa se, että aineiston tekijänoikeudellista statusta ei voida riittävän varmasti määrittää. Tekijänoikeuslainsäädännön kehittämisessä tulisikin ottaa jatkossa huomioon tutkimus- ja kulttuuriaineistojen erityiskysymykset. (Rajoituksenalaisten kieliaineistojen osalta tekijänoikeus estää erityisesti uusien aineistojen käyttöä tutkimuksessa ja opetuksessa. Lainsäädännön kehittämisen kautta on löydettävissä fair use -ratkaisuja, jotka eivät vaaranna tekijänoikeudenhaltijoiden taloudellisia intressejä mutta joilla saavutetaan merkittävää ja laaja-alaista hyötyä.)

Sanakirja-aineistojen osalta tulisi selvittää aineistojen avoimen käytön mahdollisuuksia vuodesta 2016 lähtien.

Liite 6: Korkeakoulujen metatietopääoman avaaminen

Korkeakoulujen toimintaan liittyen on tällä hetkellä tietovarantoja sekä korkeakouluilla itsellään että korkeakouluihin kohdistuvien tiedonkeruiden kautta syntyneinä eri tahoilla (OKM, OPH, Tilastokeskus jne.). Näiden tietovarantojen avaamispotentiaali tulisi selvittää.

OKM ostaa tilastokeskukselta suuren määrän koulutuspoliittista suunnittelua koskevaa aineistoa. Lähtökohtaisesti Tilastokeskuksen tulisi avata nämä aineistot kaikille avoimiksi.

OKM:n ja korkeakoulujen yhteisessä RAKETTI-hankkeessa toteutetaan HE 44/2012 mukainen korkeakoulujen valtakunnallinen tietovaranto. Tietovarantoon kootaan alkuvaiheessa säädösehdotuksen mukaisia opiskelijarekisteritietoja opiskelijavalintojen tueksi ja tiedonkeruiden tarpeisiin. Jatkossa tietovarantoa voidaan sopimus pohjalta kehittää vastaamaan tarpeisiin joissa korkeakoulujen tietoja on tarpeen koota yhteen paikkaan vastaamaan tietotarpeisiin. Tällaisia tietoja ovat esimerkiksi tutkijan henkilötiedot, joita tarvitaan tutkijan identifiointiin.

SADe ohjelman oppijan palvelukokonaisuuden osana kaikkien koulutusasteiden koulutustarjontatietoja ollaan kokoamassa Opetushallituksen koulutustiedon tietovarantoon.

Osa korkeakoulujen tiedoista (erityisesti henkilötietoja sisältämättömät tiedot) voidaan kuitenkin jakaa korkeakouluista hajautetusti ja avoimesti. Esimerkiksi Aalto-yliopisto on avannut omia tietojaan osoitteessa <http://data.aalto.fi>. Tällaisissa tapauksissa tiedon hyödynnettävyyttä edistäisi yhdenmukaisten rajapintamäärittelysten käyttäminen ja tiedon löydettävyyttä edistävät keskitetyt palvelut ja esimerkiksi tietokokonaisuuksien pysyvät URI:t. Hajautetusti avattavia tietoja voisivat olla mm. korkeakoulujen kampusten tilankäyttö, päivittäiset tapahtumakalenterit (ml. vierailut ym.) ja tutkimuksen vaikuttavuuden indikaattorit.

Keskeinen toimenpide on selvittää, miten tiedonkeruista ja muista lähteistä koottu tieto korkeakoulujen opetuksesta ja tutkimuksesta voidaan tuoda tarjolle yhtenäisellä tavalla hyödynnettäväksi. Samalla voidaan kehittää menetelmiä, joiden avulla muun muassa tieteen avoimuutta ja avointa julkaisemista voidaan kehittää. Tavoitteena on ratkaisu, jossa käyttäjät – niin kansalaiset kuin julkinen ja yksityinen sektori – löytävät tarvitsemansa tiedon helposti käyttöönsä koko korkeakoulujärjestelmästä.

Jotta keskitetyn ja hajautetun tiedon kokonaisuudesta muodostuu hallittu ja kattava kokonaisuus, tarvitaan kattava korkeakoulujen toimintaa koskevan tiedon tietoarkkitehtuuri, prosessit ja käytännöt.

Pitkällä aikavälillä tietovarantojen hyödyntämistä voidaan kehittää siten, että kerättyjen tietovarantojen pohjalta voidaan saada päätöksentekoa varten tietoa

esimerkiksi siitä, paljonko tutkimusrahoitusta on suunnattu viime vuosina päätöksenteon kannalta relevantteihin aihepiireihin ja millainen toimijoiden ekosysteemi näihin alueisiin on syntynyt. Tähän liittyvää selvitystyötä tarvitaan pikaisesti.

Liite 7: Arkistolaitoksen ns. projektitietokannat (Suomen sotasurmat 1914-1922 ym.)

Arkistolaitos tai ulkopuolinen taho on rahoittanut projekteja, joiden työn tuloksena on kerätty arkistolaitoksen aineistoista ja muodostettu tietokantoja aineiston käytön nopeuttamiseksi. Näitä ns. projektitietokantoja ovat Suomen sotasurmat 1914-1922, Suomen sodissa menehtyneet 1939-1945, Suomi, sotavangit ja ihmislouvatukset, Diplomatarium Fennicum ja Heraldica.

Tietokannoista voi arkistolaitoksen kullekin tietokannalle rakentaman ja hallinnoiman käyttöliittymän kautta hakea tietoja monilla eri tavoilla.

Tietokannat ovat dataltaan melko pieniä. Staattiset tietokannat, joihin ei tule muutoksia, voidaan muuntaa XML-muotoisiksi ja tarjota asiakkaille arkistolaitoksen kotisivulla. Kustannus on noin 5000 euroa. Dynaamisille tietokannoille voidaan luoda OAI-PMH-rajapinta. Kustannus on noin 50 000 euroa.

Hankkeen kautta avautuu mahdollisuus hyödyntää tietokannoissa olevaa tietoa uusilla tavoilla ja yhdistellä niitä esimerkiksi paikkatietokantoihin jne.

LIITE 8: KUPO:n virastojen ja laitosten tietovarantojen avaaminen

OKM:n alaisten laitosten ja virastojen tietovarannot

Yleisellä tasolla voidaan todeta, että opetus- ja kulttuuriministeriö kannattaa julkisen sektorin hallussa olevan tiedon saattamista kaikkien halukkaiden käyttöön laajasti. Tämä yleinen periaate on ilmaistu muun muassa 3.3.2011 valtioneuvoston periaatepäätöksessä julkisen sektorin digitaalisten tietoaineistojen saatavuuden parantamisesta ja uudelleenkäytön edistämisestä.

OKM:n alaisten virastojen ja laitosten tietovarannoille on tunnusomaista, että valtaosa niiden omistuksessa tai hallinnassa olevista metatieto- ja tietovarannoista on edelleen fyysisessä muodossa.

Kulttuurisektorille kuuluvien valtion virastojen ja laitosten (Kansallinen audiovisuaalinen arkisto, Museovirasto, Näkövammaisten kirjasto, Suomenlinnan hoitokunta, Valtion taidemuseo, Taiteen keskustoimikunta¹) tietovarannot ovat **sisällöltään** tyypillisesti:

- kuva-, esine-, äänite-, kirja-, äänikirja- ja liikkuvan kuvan sekä arkistoaineiston kokoelmia ja niihin liittyviä metatietovarantoja
- kulttuuriympäristön tietokantoja
- hanketietokantoja

OKM:n alaisten kulttuuri- ja kulttuuriperintölaitosten tietovarantojen tuominen laajamittaisesti tietoverkkojen kautta saataville edellyttää metatietojen tuottamista digitaaliseen muotoon sekä keskeisten fyysisessä muodossa olevien aineistojen digitointia.

Metatietovarantojen avaamisessa voidaan kulttuuri- ja kulttuuriperintölaitosten monien kokoelmien osalta edetä ripeästi, kun taas em. kokoelmia koskevien varsinaisten digitaalisten objektien avaaminen yleiseen käyttöön edellyttää avaamiseen liittyvien kysymysten selvittelyä ja/tai oikeuksien hankkimista.

Tietovarantojen käyttämistä rajoittavat **henkilötietolain** (523/1999) ja **tekijänoikeuslain** (404/1961) säännökset. Tietovarannot tai niiden osat voivat muodostaa henkilötietolaissa tarkoitettun henkilörekisterin. Tietovarantojen käsittely, ml. metatiedot, tulee järjestää henkilötietolain säännösten mukaisesti.

Tekijänoikeudella suojattujen aineistojen saattaminen yleisön saataviin esim. internetin välityksellä edellyttää tekijän tai muun oikeudenhaltijan lupaa. Oikeuksien hankinta voi tapahtua joko

1 Taiteen keskustoimikunta on 1.1.2013 alkaen Taiteen edistämiskeskus.

sopimalla suoraan oikeudenhaltijan kanssa tai kollektiivisesti asianomaisia oikeudenhaltijoita edustavien yhteishallinnointijärjestöjen kanssa.

Kulttuuri- ja kulttuuriperintölaitokset vastaavat siitä, että ne toimivat henkilötietolain ja tekijänoikeuslain säännösten mukaisesti tarjotessaan aineistoja käytettäväksi. Aineiston käyttäjä puolestaan vastaa siitä, että hän toimii lainsäädännön vaatimusten mukaisesti aineistoja käyttäessään ja niitä yleisön saataviin saattaessaan.

OKM:n kohdealueen kokonaisarkkitehtuurin merkitys tietovarantojen avaamisessa

Tietovarantojen avaamisen tulee tukeutua OKM:n kohdealueella (koulutus, tiede ja kulttuuri) rakennettavaan kokonaisarkkitehtuuriin. Sekä OKM:n hallinnoalan virastojen ja laitosten että sen hallinnonalan muiden toimijoiden tieto-, prosessi- ja järjestelmätason yhteentoimivuutta toteutetaan laaja-alaisten hankkeiden puitteissa (erityisesti KDK- ja TTA-hankkeet). Lähitulevaisuudessa merkittäviä panostuksia tarvitaan metatietojen yhdenmukaistamiseen, metatietopalvelujen käyttöönottoon, yhteisiin palveluihin, tietojärjestelmäkehitykseen ja tallennuskapasiteettiin.

Kokonaisarkkitehtuurinäkökulma on tulevissa tietovarantojen saatavuutta edistävissä kehittämishankkeissa avainasemassa, sillä tällä hetkellä virastoilla on paljon toisistaan riippumatta kehitettyjä ratkaisuja ja rajapintoja. Yhteentoimivuuden kehittäminen vaatii OKM:n kohdealueen kokonaisuuden hallintaa ja aktiivista yhteistyötä toimialan organisaatioiden kesken.

Digitaalisten kulttuurin ja tutkimuksen tietovarantojen avaaminen

Kulttuurisektorille kuuluvilla valtion virastoilla ja laitoksilla tietovarantojen ja metatietovarantojen **avaamiselle välttämättömiä toimenpiteitä** ovat:

- metatietojen täydentäminen ja päivittäminen
- datan muuttaminen koneluettavaan muotoon
- palvelurajapintojen suunnittelu ja toteutus
- metatietovarantojen ja tietovarantojen latauspalvelun toteuttaminen
- yhteentoimivuuden edistäminen OKM:n kohdealueella

Kulttuuri- ja kulttuuriperintöaineistojen **avaamisessa huomioonotettavia keskeisiä kysymyksiä** ovat näyttö- ja

käyttörajoitteisten aineistojen käytön edellyttämien kansallisten palveluiden kehittäminen ja jatkuva ylläpito sekä integroitavuus palvelukerrokseen. Kulttuuriaineistojen rajatun käytön edistämiseksi keskeistä on aineistojen **etäkäyttösovellusten** kehittäminen sekä **tunnistautumispalvelut**.

Esiselvitysvaiheessa olevan kansallisen metatietopalvelun tunnistetusta rakenneosista etenkin **ontologiapalvelun, tunniste palvelun ja nimitietopalvelun** ylläpito ja kehittäminen ovat tietovarantojen yhteentoimivuuden ja avaamisen näkökulmasta keskeisessä roolissa.

Kansallinen digitaalinen kirjasto ja Tutkimuksen tietoaaineistot -hankkeet toteuttavat tietovarantojen saatavuutta

Digitaalisten kulttuuriperintöaineistojen ja tutkimuksen tietoaaineistojen hallinnan, jakelun ja säilyttämisen tietoinfrastruktuuria kehitetään Kansallinen digitaalinen kirjasto -hankkeessa (KDK) ja Tutkimuksen tietoaaineistot -hankkeessa (TTA).

Kulttuurisektorin toimijoiden tietovarantojen saatavuutta edistetään aktiivisesti Kansallinen digitaalinen kirjasto -hankekokonaisuudessa (KDK), jolla tuodaan kirjastojen, arkistojen ja museoiden digitaaliset aineistot ja palvelut helposti saataville (KDK-asiakasliittymä) ja varmistetaan kulttuuriperintöaineistojen luotettava pitkäaikaissäilytys (KDK-PAS). KDK- ja TTA-hankkeista on toimitettu TTS-prosessiin erillinen muistio (LIITE 2).

Asiakasliittymä ja datakatalogi

Kansallinen digitaalinen kirjasto -hankekokonaisuudessa otetaan vuonna 2013 käyttöön **asiakasliittymä kirjastojen, arkistojen ja museoiden aineistoihin ja palveluihin**. Asiakasliittymä mahdollistaa metatiedon jatkojalostamisen mm. sosiaalisella metatiedolla sekä metatiedon aggregoinnin Europeana-portaaliin (Euroopan digitaalinen kirjasto). Europeanan kautta avattiin 2012 yli 20 miljoonan kohteen, ml. **700 000 suomalaisen kulttuuriperintöobjektin, metatiedot vapaasti hyödynnettäväksi**.

KDK-hanke ja TTA-hanke kehittävät ja ottavat käyttöön yhteisiä palveluja, joilla edistetään digitaalisten tietovarantojen hallintaa, saatavuutta ja pitkäaikaissäilytystä. TTA-hankkeessa toteutettava datakatalogi tutkimusaineistoille pohjautuu geneeriseen, avoimen lähdekoodin ohjelmistoon, jota on tarkoituksenmukaista hyödyntää

myös kehitettäessä latauspalvelu, jolla museoiden, kirjastojen ja arkistojen tietoaaineistoja voidaan välittää koneluettavassa muodossa.

Digitaalisten kulttuuriperintöaineistojen ja tutkimuksen tietoaaineistojen pitkäaikaissäilytysjärjestelmä

Digitaalisten kulttuuriperintöaineistojen ja tutkimuksen tietoaaineistojen pitkäaikaissäilytys-järjestelmän toteuttaminen Kansallinen digitaalinen kirjasto -hankkeessa (KDK) sekä Tutkimuksen tietoaaineistot -hankkeessa (TTA) valmistettava pitkäaikaissäilytysjärjestelmä (KDK-PAS ja TTA-PAS) vähentävät digitaalisten kulttuuriperintöaineistojen ja tutkimuksen tietoaaineistojen hallinnan, jakelun ja säilyttämisen päällekkäisiä toimia ja pidemmällä aikavälillä kustannusten nousua. Yhteinen digitaalisten aineistojen pitkäaikaissäilyttämisen ja saatavuuden ratkaisu tuottaa korkealaatuisia palveluita kustannustehokkaasti ja mahdollistaa digitaalisten aineistojen monipuolisen hyödyntämisen.

Tutkimuksen tietoaaineistojen pitkäaikaissäilytyksen ratkaisu (TTA-PAS) on välttämätön, jotta voidaan turvata tulevien tutkijasukupolvien käyttöön tutkimuksen digitalisoitumisen myötä yhä nopeammin kasvavat tietomäärät. Eri tieteenaloilla on jo nyt erittäin arvokkaita tietoaaineistoja, esimerkiksi vuosikymmenien aikana kerättyjä aikasarjoja (ilmastontiedot, koulujen oppimistulokset).

Opetus- ja kulttuuriministeriön hallinnonalalla on lukuisia toimijoita, joilla on lakisääteinen velvoite tai muu perusteltu tarve säilyttää digitaalista aineistoa pysyvästi tai erittäin pitkäaikaisesti, mutta nykyiset tietojärjestelmät eivät sisällä pitkäaikaissäilytyksen edellyttämiä toiminnallisuuksia. On erityisen tärkeää, että digitaalisten kulttuuri- ja tutkimusaineistojen säilytys turvataan tuleville sukupolville.

Kulttuurisektorin virastojen ja laitosten tietovarantojen avaaminen aineistotyypeittäin 2013 - 2017

- Yhteinen latauspalvelu kulttuurisektorin virastojen ja laitosten tietovarantojen avaamiseksi koneluettavassa muodossa: Latauspalvelun toteuttaminen on välttämätön edellytys aineistojen avaamiselle
 - 300 000 euroa / 2013 - 2014.
- Kulttuuriympäristön tietovarannot (Museovirasto) sekä Suomenlinnan hoitokunnan tietovarannot. Avattu osin (Museoviraston latauspalvelu toiminnassa ; Paikkatietoikkuna / karttaikkuna) (tyyppi 1*)
 - Tietovarantojen kertaluonteinen avaaminen 2013: 100 000.
 - Sisältöjen laajentaminen 2014 - 2016: 250 000 / vuosi.
- Esinekokoelmien metatiedot ja osa digitaalisista objekteista (Museovirasto) sekä taideteosten metatiedot ja osa digitaalisista objekteista (Valtion taidemuseo). Avattu osin. (tyyppi 1, 2 ja 3).
 - Sisältöjen laajentaminen 2013 - 2017. 450 000 euroa v. / 2014 - 2015 ja 350 000 euroa v. / 2016 - 2017.
- Kirjastojen metatietovarannot / Näkövammaisten kirjasto, Museoviraston kirjasto ja Valtion taidemuseon kirjasto. Museoviraston kirjaston metatietovaranto avataan 2013. Muut avattavissa 2014 - 2016. (tyyppi 1 ja 2).
 - 50 000 v. / 2014 - 2015.
- Kansatieteen ja historian kuvakokoelmat (Museovirasto) sekä kuvataiteen keskusarkiston valokuvakokoelmat, kuvataiteen ja taidehistorian arkistotietokannat ja AV-tietokanta (Valtion taidemuseo). Avattu osin (tyyppi 1, 2 ja 3).
 - Sisältöjen laajentaminen ja avaaminen vaiheittain 2014 - 2017. 250 000 euroa v. / 2014 - 2017.
- Elokuvien filmografiset tiedot (Kansallinen audiovisuaalinen keskus KAVA). Tyyppi 2.
 - Avattavissa 2013: 20 000 euroa.
- Apuraha- ja määräraharekisteri sekä hankerekisteri (Taiteen keskustoimikunta, 1.1.2013 alkaen Taiteen edistämiskeskus.) (tyyppi 2 ja 3).
 - Avattavissa 2013 - 2014.

***Aineistotyyppien selitteet:**

- 1) jo avattu (saatavissa koneluettavassa muodossa maksutta)
- 2) julkinen, avattavissa oleva, johon ei liity mitään käyttörajoituksia tai lainsäädännön esteitä avaamiselle
- 3) tietovaranto, joka mahdollisesti voitaisiin ja voisi olla hyödyllistä avata, mutta jonka avaamiseen liittyvät kysymykset (esim. tekijänoikeus) vaativat erillistä selvittelyä ja oikeuksien hankkimista

4) salassa pidettävä, joka voidaan jättää tämän selvityksen ulkopuolelle.

LIITE 9

TTS 2014 - 2017

Momentit
29.40.02
29.40.20
29.40.50
29.80.55

Digitaalisten kulttuuriperintöaineistojen ja tutkimuksen tietoaaineistojen pitkäaikaissäilytysjärjestelmän toteuttaminen

Kansallinen digitaalinen kirjasto -hankkeessa (KDK) ja Tutkimuksen tietoaaineistot -hankkeessa (TTA) valmisteltava pitkäaikaissäilytysjärjestelmä (KDK-PAS ja TTA-PAS) vähentää digitaalisten kulttuuriperintöaineistojen ja tutkimuksen tietoaaineistojen hallinnan, jakelun ja säilyttämisen päällekkäisiä toimia ja pidemmällä aikavälillä kustannusten nousua. **Yhteinen digitaalisten aineistojen pitkäaikaissäilyttämisen ja saatavuuden ratkaisu tuottaa korkealaatuisia palveluita kustannustehokkaasti ja mahdollistaa digitaalisten aineistojen monipuolisen hyödyntämisen.** Hanke edistää OKM:n kohdealueen (koulutus, tiede ja kulttuuri) yhteentoimivuutta ja tiedon yhteiskäyttöä yhtenäistämällä kuvauksia sekä kehittämällä yhteisiä säilytyksen ja saatavuuden ratkaisuja.

Kansallinen digitaalinen kirjasto -hankkeessa tuodaan kirjastojen, arkistojen ja museoiden digitaaliset aineistot ja palvelut helposti kansalaisten saataville (KDK-asiakasliittymä) ja varmistetaan kulttuuriperintöaineistojen luotettava pitkäaikaissäilytys (KDK-PAS). TTA-hankkeessa kehitetään ratkaisu tutkimusaineistojen ja niiden metatietojen tallentamiseen sekä aineistojen säilyttämiseen. Tutkimuksen tietoaaineistojen pitkäaikaissäilytyksen ratkaisu (TTA-PAS) on välttämätön, jotta voidaan turvata tulevien tutkijasukupolvien käyttöön tutkimuksen digitalisoitumisen myötä yhä nopeammin kasvavat tietomäärät. Eri tieteenaloilla on jo nyt erittäin arvokkaita tietoaaineistoja, esimerkiksi vuosikymmenien aikana kerättyjä aikasarjoja (ilmastontiedot, koulujen oppimistulokset).

Hankekokonaisuus toteuttaa hallitusohjelman kirjauksia julkisten tietovarantojen avaamisesta ja hyödyntämisestä, kansallisen tutkimusinfrastruktuuripolitiikan edistämisestä sekä kirjastojen, museoiden ja julkisten taidelaitosten digitaalisten aineistojen avaamisesta. Hanke on kirjattu toimenpiteenä Valtioneuvoston periaatepäätökseen tietoyhteiskuntapolitiikan tavoitteista sekä Valtioneuvoston periaatepäätökseen julkisen sektorin digitaalisten tietoaaineistojen saatavuuden parantamisesta ja uudelleenkäytön edistämisestä. Ministeriön hallinnonalalla on lukuisia toimijoita, joilla on lakisääteinen velvoite tai muu perusteltu tarve säilyttää digitaalista aineistoa pysyvästi tai erittäin pitkäaikaisesti, mutta nykyiset tietojärjestelmät eivät

sisällä pitkäaikaissäilytyksen edellyttämiä toiminnallisuuksia. **On erityisen tärkeää, että digitaalisten kulttuuri- ja tutkimusaineistojen säilytys turvataan tuleville sukupolville.**

<i>Esitetään seuraavaa lisämäärärahan tarvetta</i>	<i>2014</i>	<i>2015</i>	<i>2016</i>	<i>2017</i>
KDK ja TTA yhteensä (1000 euroa)	940	740	930	860
	0	0	0	0
KTPO:n osuus	7900	5900	6800	6100
TTA-PAS (Tutkimuksen tietoaaineistojen pitkäaikaissäilytys) (29.40.20)	4 200	3 400	2 500	2 500
KDK-PAS (Digitaalisten kulttuuriperintöaineistojen pitkäaikaissäilytys (29.40.20)	2 800	1 600	3 400	2600
KDK:n asiakasliittymä (29.40.50)	500	500	500	500
Sähköisten aineistojen vastaanotto- ja palvelujärjestelmä VAPA (29.40.02)	400	400	400	400
KUPO:n osuus	1 500	1 500	2 500	2500
KDK-PAS ja asiakasliittymä (29.80.55)	1 500	1 500	2 500	2500

Liite 10

Koulutuspolitiikan osaston toimialan suunnitelmat

Tietovarannon nimi: **Koulutustoimialan koodistot**

Tietovarannon vastuuorganisaatio: Opetushallitus

Tietovarannon sisällön kuvaus lyhyesti:

- koulutustoimialan koodit ja muut koodit koulutustoimialan käyttöön
- koulutustoimialan koodit kaikkien käytettäväksi
- koodit kuvattu osoitteessa:
<https://confluence.csc.fi/display/oppija/Koodistot+ja+koodit>

Tietovarannon säädösperusta

- ei tarvita

Tietojen luovuttamisesta tai tietosuoritteista nykyisin perittävät maksut ja niiden peruste

- ilmainen

Tietovaraston avaamisen ajankohta (jo toteutunut tai suunniteltu)

- 01/2014

Tarvittavat toimenpiteet tietovarannon avaamiseksi koneluettavassa muodossa ja maksutta

- toteutetaan osana SADe hanketta

Arvio tietovarannon avaamisen taloudellisista vaikutuksista (kustannuksista ja hyödyistä)

- päällekkäisyyksien poistaminen

Tietovarannon avaamisen vaikutukset lainsäädäntöön ja sen kehittämistarpeisiin

- ei vaikutusta

Tietovarannon nimi: **Koulutustarjonta**

Tietovarannon vastuujärjestelmä: Opetushallitus

Tietovarannon sisällön kuvaus lyhyesti:

- koulutustarjonnan tietomalli- ja tietosisältö kuvattu osoitteessa <https://confluence.csc.fi/display/kshj/Arkkitehtuuri>

Tietovarannon säädöspohja:

- Luodaan osana SADe hanketta

Tietojen luovuttamisesta tai tietosuoritteista nykyisin perittävät maksut ja niiden peruste:

- ilmainen

Tietovarannon avaamisen ajankohta (jo toteutunut tai suunniteltu):

- 01/2014

Tarvittavat toimenpiteet tietovarannon avaamiseksi koneluettavassa muodossa ja maksutta:

- ei

Arvio tietovarannon avaamisen taloudellisista vaikutuksista (kustannuksista ja hyödyistä)

- Avoin rajapinta toteutetaan osana SADe hanketta
- Tietoja voidaan käyttää osana liiketoimintaa
- Päällekkäisen tiedon keräämisen vähentäminen

Tietovarannon avaamisen vaikutukset lainsäädäntöön ja sen kehittämistarpeisiin

- Lainsäädännön muutokset osana SADe hanketta

Tietovarannon nimi

Näyttötutkintojen palautejärjestelmä AIPAL

Tietovarannon vastuuorganisaatio

Opetushallitus

Tietovarannon sisällön kuvaus lyhyesti

Näyttötutkintojen suorittajien palautetta näyttötutkintojärjestelmän toimivuudesta liittyen hakeutumiseen, tutkinnon suorittamiseen ja tarvittavan ammattitaidon hankkimiseen. Palaute annetaan anonyyminä. Valtakunnalliset tutkinnon järjestäjä-, tutkinto- ja opintoalakohtaiset tulokset ovat tilastomuodossa. Tutkinnon järjestäjäkohtaiset tulokset, jotka saadaan heidän omien palautekysymystensä kautta, eivät ole julkisia.

Tietovarannon säädösperusta

Laki ammatillisesta aikuiskoulutuksesta 1998/631 8 § 3 mom, jonka perusteella OPH:n määräys 43/11/2006 1 §.

Tietojen luovuttamisesta tai tietosuoritteista nykyisin perittävät maksut ja niiden peruste

Ilmainen.

Tietovaraston avaamisen ajankohta (jo toteutunut tai suunniteltu)

Järjestelmää ollaan kehittämässä 2013, jolloin on mahdollista rakentaa rajapintoja julkisiin tietoihin.

Tarvittavat toimenpiteet tietovarannon avaamiseksi koneluettavassa muodossa ja maksutta

Arvioitava erikseen.

Arvio tietovarannon avaamisen taloudellisista vaikutuksista (kustannuksista ja hyödyistä)

Kyseessä on laadunvarmistusjärjestelmä, jolla ei ole välittömiä taloudellisia vaikutuksia.

Tietovarannon avaamisen vaikutukset lainsäädäntöön ja sen kehittämistarpeisiin

Ei vaikutuksia.

Tietovarannon nimi:

Ammatillisen koulutuksen, perusopetuksen ja lukiokoulutuksen opetussuunnitelmien ja tutkintojen perusteet

Tietovarannon vastuuorganisaatio: Opetushallitus

Tietovarannon sisällön kuvaus lyhyesti:

- Tietosisältö kuvattu osoitteessa <https://confluence.csc.fi/display/kshj/Arkkitehtuuri>
- mm. opintokokonaisuudet, tutkinnon osat ja osaamisen vaatimukset

Tietovarannon säädösperusta:

- ei lainsäädäntö muutoksia

- rekisteriseloste luotava

Tietojen luovuttamisesta tai tietosuoritteista nykyisin perittävät maksut ja niiden peruste

- ilmainen

Tietovaraston avaamisen ajankohta (jo toteutunut tai suunniteltu)

- 09/2015

Tarvittavat toimenpiteet tietovarannon avaamiseksi koneluettavassa muodossa ja maksutta

- Toteutetaan osana SADe hanketta

Arvio tietovarannon avaamisen taloudellisista vaikutuksista (kustannuksista ja hyödyistä):

- Antaa mahdollisuuden kehittää uusia liiketoimintaratkaisuja tietoa hyödyntäen
- integroituu samattomasti osaksi todennetun osaamisen rekisteriä

Tietovarannon avaamisen vaikutukset lainsäädäntöön ja sen kehittämistarpeisiin

- ei vaikutuksia lainsäädäntöön (rekisteri ei sisällä henkilötietoja)

Tietovarannon nimi: **Koulutustoimialan organisaatiotiedot**

Tietovarannon vastuuorganisaatio: Opetushallitus

Tietovarannon sisällön kuvaus lyhyesti

Tietovarannossa:

- koulutuksen järjestäjät yhteystietoineen
- opetuksen järjestäjät yhteystietoineen
- oppilaitokset ja toimipisteet yhteys- ja kuvaustietoineen

Tietovarannon säädösperusta

- Luodaan osana SADe hanketta

Tietojen luovuttamisesta tai tietosuoritteista nykyisin perittävät maksut ja niiden peruste

- Ilmainen

Tietovaraston avaamisen ajankohta (jo toteutunut tai suunniteltu)

- Avoin rajapinta avataan 03/2014

Tarvittavat toimenpiteet tietovarannon avaamiseksi koneluettavassa muodossa ja maksutta

- ei toimenpiteitä

Arvio tietovarannon avaamisen taloudellisista vaikutuksista (kustannuksista ja hyödyistä)

- Avoin rajapinta toteutetaan osana SADe hanketta
- Tietoja voidaan käyttää osana liiketoimintaa
- Päällekkäisen tiedon keräämisen vähentäminen

Tietovarannon avaamisen vaikutukset lainsäädäntöön ja sen kehittämistarpeisiin

- Lainsäädännön muutokset osana SADe hanketta

Tietovarannon nimi: **Todennetun osaamisen rekisteri**

Tietovarannon vastuuorganisaatio: Opetushallitus

Tietovarannon sisällön kuvaus lyhyesti:

- Perusopetuksen päättötodistusten suoritustiedot ja arvosanat
- Lukion päättötodistus ja suoritustiedot
- Perusopetuksen lisäopetuksen päättötodistus ja suoritustiedot
- Ammattistarttin suoritustodistus

- Kotitalousopetuksen suoritustodistus
- Todistus kansanopistolinjan suorittamisesta (vähintään 1100 tuntia tai 20 ov)
- ammatillisen perustutkinnon päättötodistus ja suoritustiedot
- ammatti- ja erikoisammattitutkintojen päättötodistus ja suoritustiedot
- VIRTAtietovarannosta siirrettävät korkeakoulujen todistus ja opintosuoritustiedot

Tietovarannon säädösperusta:

- Lakivalmistelu eduskunnassa korkeakoulujen osalta
- Luodaan osana SADe hanketta muilta osin

Tietojen luovuttamisesta tai tietosuoritteista nykyisin perittävät maksut ja niiden peruste

- ilmainen

Tietovaraston avaamisen ajankohta (jo toteutunut tai suunniteltu)

- 01/2015

Tarvittavat toimenpiteet tietovarannon avaamiseksi koneluettavassa muodossa ja maksutta

- Toteutetaan osana SADe hanketta

Arvio tietovarannon avaamisen taloudellisista vaikutuksista (kustannuksista ja hyödyistä)

- Arvio tehty osana SADe hanketta

Tietovarannon avaamisen vaikutukset lainsäädäntöön ja sen kehittämistarpeisiin

- Lainsäädäntö muutokset osana SADe hanketta sekä korkeakoulujen VIRTAtietovarannosta siirrettävät korkeakoulujen todistus ja opintosuoritustiedot

Resurssitarve

- 75000e/vuosi (1 HTV) - ei sisälly Opetushallituksen TTS 2014 - 2017 esitykseen.

Tietovarannon nimi

Tutkintotoimikuntarekisteri

Tietovarannon vastuorganisaatio

Opetushallitus

Tietovarannon sisällön kuvaus lyhyesti

Tutkintotoimikuntien toimialat ja jäsenet sekä näyttötutkintojen järjestämissopimustiedot.

Tietovarannon säädöspöerusta

Laki ammatillisesta aikuiskoulutuksesta 1998/631 ja Asetus ammatillisesta aikuiskoulutuksesta 1998/812.

Tietojen luovuttamisesta tai tietosuoritteista nykyisin perittävät maksut ja niiden peruste

Ilmainen.

Tietovaraston avaamisen ajankohta (jo toteutunut tai suunniteltu)

Järjestelmää ollaan uudistamassa 2013, jolloin on mahdollista rakentaa rajapintoja julkisiin tietoihin.

Tarvittavat toimenpiteet tietovarannon avaamiseksi koneluettavassa muodossa ja maksutta

Arvioitava erikseen.

Arvio tietovarannon avaamisen taloudellisista vaikutuksista (kustannuksista ja hyödyistä)

Näyttötutkintojen järjestäjätietojen saatavuudella poistetaan päällekkäistä työtä.

Tietovarannon avaamisen vaikutukset lainsäädäntöön ja sen kehittämistarpeisiin

Ei vaikutuksia.

Tietovarannon nimi: **Opetus- ja kulttuuritoimen valtionosuudet, valtionosuussuoritteet ja taloustiedot**

Tietovarannon vastuorganisaatio: Opetushallitus minusta vastuorganisaatio on OKM

Tietovarannon sisällön kuvaus lyhyesti:

- Opetus- ja kulttuuritoimen valtionosuudet ja niiden laskennassa käytettävät laskentatiedot sekä niihin liittyviä tilastotietoja (opiskelija- tai muita suoritelmääriä, kustannustietoja ja yksikköhintoja, tietoja rahoituksen määrästä ja maksatuksista) kaikista koulutusmuodoista (pl. yliopistot ja opintokeskukset) ja kaikilta kulttuurihallinnon sektoreilta

Tietovarannon säädösperusta:

- Laki ja asetus opetustoimen ja kulttuuritoimen rahoituksesta
- Laki ja asetus vapaasta sivistystyöstä
- Museolaki
- Teatteri- ja orkesterilaki

Tietojen luovuttamisesta tai tietosuoritteista nykyisin perittävät maksut ja niiden peruste

- Tietoa julkaistaan konekielisinä raportteina (Excel- ja/tai html-muodossa) ilmaiseksi. Lisäksi on mahdollista tuottaa itse monia konekielisiä raportteja suoraan tietokannasta. Muun tietoaineiston luovutuksesta Tampereen yliopisto laskuttaa aiheutuneen työmäärän mukaan.

Tietovaraston avaamisen ajankohta (jo toteutunut tai suunniteltu)

- ensimmäiset tiedot ovat vuodelta 1995 , josta saatavien tietojen määrä on koko ajan kasvanut ja kattaa nyt kaikki osa-alueet ja sektorit
- Järjestelmän uudistamisesta ei ole päätöstä. ??? Ei tietovaraston avaminen vaadi järjestelmän uudistamista. Tarvitsee vain määritellä rajapinnat ja toteuttaa ne, mutta ne pitää myös ylläpitää, koska järjestelmän sisältämä tieto karttuu joka vuosi eikä yleensä aivan edellisen vuoden mukaisena
- rajapintaa voidaan tarjota aikaisintaan vuonna 2017 (tarve arvioitava kuitenkin ennen julkaisua). Teknistä estettä rajapintojen tarjoamiseksi vaikka vuonna jo 2014 ei ole, kunhan määritellään rajapinta ja siinä tarjottavat tiedot.

Tarvittavat toimenpiteet tietovarannon avaamiseksi koneluettavassa muodossa ja maksutta

- järjestelmässä olevat tiedot ovat jo nyt pääosin yleisesti konekielisesti luettavissa internetissä Excel-, html-tiedostoina ja/tai muuten konekielisesti helposti luettavina raportteina maksutta.

Arvio tietovarannon avaamisen taloudellisista vaikutuksista (kustannuksista ja hyödyistä)

- arvioitava osana järjestelmäkokonaisuuden uudistamisen esitutkimusta (kuka on tekemässä ja/tai koska ja missä tekeminen on tarkoitus tehdä ja koska aloittaa?)

Tietovarannon avaamisen vaikutukset lainsäädäntöön ja sen kehittämistarpeisiin

- arvioitava osana järjestelmäkokonaisuuden uudistamisen esitutkimusta