


## TUTKIMUKSEN TUOTOSTEN PITKÄAIKAISAAVUUDEN TOIMINTAMALLI

Pirjo-Leena Forsström, 16.12.2015, versio 1.0

<b>Toimintamalliin vaikuttaa</b>	<ul style="list-style-type: none"> <li>- Tutkimuksen viitearkkitehtuuri</li> <li>- Yhteentoimivuusmalli</li> <li>- JHKA KDK standardisalkku <a href="http://www.kdk.fi/images/tiedostot/KDK_standardisalkku27_10_2014.pdf">http://www.kdk.fi/images/tiedostot/KDK_standardisalkku27_10_2014.pdf</a></li> </ul>
<b>Toimintamalli vaikuttaa</b>	<ul style="list-style-type: none"> <li>- Tutkimus-PAS vaatimusmäärittelyt</li> <li>- PAS-ratkaisun kehityssuunnitelma</li> <li>- PAS-pilotointisuunnitelma</li> </ul>

<b>Julkaisu</b> Tutkimuksen tuotosten pitkäaikaissaatavuuden toimintamalli	
<b>Julkaisija</b> Avoin tiede ja tutkimus -hanke	<b>Julkaisuajankohta</b> 14.12.2015
<b>Tekijä</b> Pirjo-Leena Forsström	
<b>Lisenssi</b>	
 <p>Tämä teos on lisensoitu <a href="https://creativecommons.org/licenses/by/4.0/">Creative Commons Nimeä 4.0 Kansainvälinen -lisenssillä</a>.</p>	
<b>Julkaisun jakelu</b> PDF-tiedosto ladattavissa <a href="http://avointiede.fi/keskeiset-julkaisut">avointiede.fi/keskeiset-julkaisut</a> <a href="http://urn.fi/URN:NBN:fi-fe2016122731711">http://urn.fi/URN:NBN:fi-fe2016122731711</a>	
<b>Yhteystiedot</b> <a href="http://avointiede.fi">http://avointiede.fi</a> <a href="mailto:avointiede@postit.csc.fi">avointiede@postit.csc.fi</a>	



## SISÄLTÖ

Tutkimuksen tuotosten pitkäaikaissaatavuuden toimintamalli .....	1
Yhteenveto .....	3
Käsitteet .....	3
Tavoitteet ja tausta .....	4
Toimintamalli .....	5
Toimijat.....	8
Sopimukset.....	8
Resursointipäätökset.....	9
Kustannusten hallintamalli .....	9
Toimintamallin kehittäminen .....	9
Tunnistettuja tutkimusaineistoja.....	10
Toimintamallin arkkitehtuuri .....	11
Prosessien kokoaminen .....	12
Prosessien yhteentoimivuus .....	12
Toimintamallin prosessit.....	13
Alustava yhteisen palveluarkkitehtuurin malli .....	15
Yhteenveto .....	15

## YHTEENVETO

Tutkimuksen tulokset ovat kansallisesti merkittävä tutkimusaineistokokonaisuus, joiden laajan käytön ja luotettavan saatavuuden takaaminen on suomalaisen tutkimus- ja innovaatio toiminnan ydin. Tutkimuksen tuloksien tärkein ero ja piirre esimerkiksi kulttuuriperintöaineistoihin nähden on pitkäaikaissaatavuuden tarkoitus: säilytetään, jotta voidaan hyödyntää uuden tiedon tuottamisessa. Tieteen ja tutkimuksen perusta on kestävä, kun tutkimuksen tuottamien tuotosten pitkäaikaissaatavuus on toteutettu kustannustehokkaasti. Tällöin tutkimustuloksia voidaan sekä hyödyntää pidempään että turvata tulevien tutkijasukupolvien käyttöön tutkimuksen digitalisoitumisen myötä yhä nopeammin kasvavat tietomäärät.

Tutkimuksen tuotosten pitkäaikaissaatavuuden turvaamisen toimintamallilla tarkoitetaan suunnitelmaa seuraavien 5-10 vuoden aikana tapahtuvasta uudistamistyöstä. Pitkäaikaissaatavuutta tukevan kokonaisuuden tulee mahdollistaa ja tukea eri tieteenalojen ja toimintojen tarpeita, sekä olla muokattavissa oleva kerroksellinen, modulaarinen ja kustannustehokas ratkaisumalli, jossa hyödynnetään olemassa olevia komponentteja. Kokonaisuuden tulee mahdollistaa tutkimuksen tuotosten avoimuus ja jatkokäyttö primäärilähteistään.

Toimintamallin tulee olla:

- **Läpinäkyvä toimijoille.** Läpinäkyvyys on keino toimijoiden luottamukseen ja sitoutumiseen.
- **Monimuotoinen ja hajautettua arkkitehtuuria tukeva.** Tällöin laaja toimijakentän osaaminen ja tarpeet voidaan parhaiten hyödyntää.
- **Kustannustehokas ja taloudellisesti kestävä.**
- **Erilaisia liiketoimintamalleja tukeva** (esim. kansainvälisessä yhteistyössä).
- **Erilaisia exit-ratkaisuja tukeva.** Kansalliselta alueelliselle, kansalliselle ja kansainväliselle tasolle siirryttäessä materiaalien siirto palvelusta toiseen tulee olla helppoa.

Toimintamalli kytkeytyy korkeakoulujen sisäiseen toiminnan kokonaisuuteen, mutta myös kansainvälisiin tutkimusinfrastruktuureihin. Erityisesti yhteisiä kansallisia palveluita tulee kehittää näiden kytköksen kautta.

## KÄSITTEET

**Tutkimustuotos** on se konkreettinen väline, johon tutkimuksen sisältö ja tutkimustulos kuvataan ja jolla tulos ilmaistaan.

**Tutkimusaineisto** on tutkimuksessa tuotettava ja/tai hyödynnettävä materiaali jota analysoimalla pyritään vastaamaan tutkimusongelmaan.

**Pitkäaikaissaatavuus** tarkoittaa digitaalisen informaation luotettavaa tarjoamista käyttöön useiden kymmenien tai jopa satojen vuosien ajan. Saatavuuden turvaaminen koostuu säilytyksestä eri tasoinen sekä käyttökelpoisuuden varmistamisesta.

## TAVOITTEET JA TAUSTA

Digitaalisen informaation pitkäaikaisen saatavuuden turvaaminen on sellaisten kestävien toimintamallien rakentamista, joilla voidaan varmistaa aineistojen käyttökelpoisuus. Tarkoituksenmukaisen toimintatavan valinta perustuu läpinäkyvään kustannusten ja hyötyjen tasapainottamiseen pitkällä aikavälillä. Saatavuuden turvaaminen koostuu säilytyksestä eri tasoinen sekä käyttökelpoisuuden varmistamisesta. Eri tieteenaloilla on erittäin arvokkaita tutkimusaineistoja, kuten vuosikymmenien aikana kerättyjä aikasarjoja (esimerkiksi ilmastotiedot, koulujen oppimistulokset). Tutkimuksen tuloksien tärkein ero ja piirre esimerkiksi kulttuuriperintöaineistoihin nähden on pitkäaikaissaatavuuden tarkoitus: säilytetään, jotta voidaan hyödyntää uuden tiedon tuottamisessa.

Tutkimusprosessin perusta ovat osaavat ihmiset sekä heidän kauttaan tutkimuksen tuotosten elinkaaren ja laadun hyvään suunnitteluun ja hallintaan. Tieteen ja tutkimuksen perusta on kestävä, kun tulokset ovat laajasti saatavilla sekä luotettavasti säilytettyjä. Validointia, verifiointia ja toistettavuutta varten täytyy seuraavien tutkimusmateriaalien (eli tutkimusaineistojen ja tutkimustuotosten) olla suunnitellusti, pitkäaikaisesti ja yhtee toimivasti saatavilla.

ATT-Viitearkkitehtuuri luo kokonaiskuva avoimen tieteen ja tutkimuksen osa-alueista ja sitä tukevista tietojärjestelmistä. Avoimen tieteen ja tutkimuksen viitearkkitehtuurilla muodostetaan tavoitetila, joka mahdollistaa nykyisen tutkimusprosessin kaikkien vaiheiden avoimuuden ja jolla voidaan kiihdyttää tieteen tekemistä, mutta myös mahdollistaa uudenlaisen, ketterämmän ja yhteisöllisemmän tavan tutkia ja tehdä tiedettä.

### Tutkimuksen hyödyntäminen

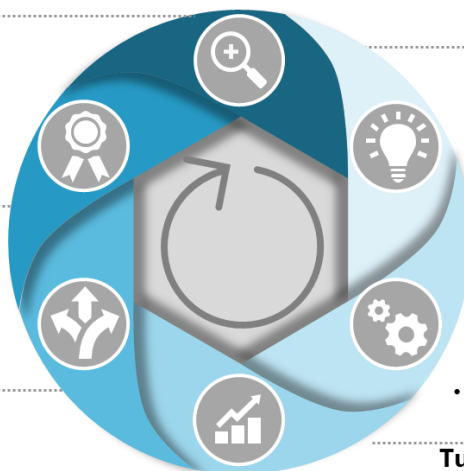
- Tutkimustulosten hyödyntäminen
- Tutkimustuotosten hyödyntäminen
- Mittaaminen
- Tutkimustuotosten kuratointi

### Tulosten arvottaminen

- Tutkimustulosten arvottaminen
- Tuotosten säilyttämistarpeen arviointi

### Julkaisu ja levittäminen

- Tutkimustuotosten arviointi ja julkaisu
- Tutkimustuotosten jakelu
- Tutkimustulosten viestintä
- Pitkäaikaissaatavuuden varmistaminen



### Tutkimuksen valmistelu

- Tutkimusongelman määrittäminen
  - Tutkimuksen suunnittelu
- Tutkimusympäristön määrittäminen
- Tutkimuksen resursointi ja rahoitus
  - Tutkimuksen organisointi

### Tutkimuksen toteutus

- Tutkimusmenetelmän valinta tai kehittäminen
- Tutkimusaineiston valinta ja kokoaminen
    - Tutkimusaineiston käsittely
  - Tutkimustapahtumien dokumentointi
  - Tutkimustulosten löytäminen & tunnistaminen

### Tutkimustuotosten tuottaminen

- Tutkimustuotosten tuottaminen, tutkimustulosten kuvaaminen
  - Tutkimustuotosten ja menetelmien tallennus ja säilytys

Tavoitearkkitehtuuriin kuuluvia tärkeimpiä linjauksia ovat:

- Tieteen avoimuutta edistetään kaikissa prosessin vaiheissa
- Tutkimusaineistot ja tuotokset ovat lähtökohtaisesti avoimina tiedeyhteisölle
- Tutkimusaineistoja käytetään metahaun kautta keskitetysti

Tutkimuksen tuotosten pitkäaikaissaatavuuden turvaamisen toimintamallilla tarkoitetaan suunnitelmaa seuraavien 5-10 vuoden aikana tapahtuvasta uudistamistyöstä. Kun pitkäaikaissaatavuuden tavoite on saavutettu:

- Hyvin toteutettu pitkäaikaissaatavuuspalvelu on tutkijoille helppo ja kannustava tapa toimia, ja aineistojen elinkaaresta huolehtiminen on luonteva osa tutkimuksen käytäntöjä.
- Tutkimusdatat, tutkimusjulkaisut ja tutkimusmenetelmät ovat helposti löydettävissä ja saatavilla. Vastuut näiden elinkaaren eri vaiheisiin liittyvistä tehtävistä on sovittu ja kaikkien tiedossa, ja tehtävät on resursoitu riittävästi.
- Avoin tietoaarkkitehtuuri mahdollistaa yhteentoimivan tutkimustietoinfrastruktuurin kehittämisen. Pitkäaikaissaatavuuteen liittyvä standardointityö on edistänyt tutkimusinfrastruktuurien ja korkeakoulujen parempaa tutkimustiedon hallintaa. Arkistot, kirjastot ja museot ovat luonteva osa tätä kokonaisuutta.
- Perusta on kunnossa suuriin haasteisiin vastaamiseksi, ja tutkimustieto ja tutkimusmenetelmät ovat aktiivisessa käytössä. Palvelut käytön helpottamiseksi ovat helposti löydettävissä ja niitä kehitetään toimintalähtöisesti.
- Toiminnan yhteentoimivuus kansallisesti ja kansainvälisesti on hallittua, selkeästi kuvattua ja helppoa.

Pitkäaikaissaatavuudeltaan turvattavat tutkimuksen tulokset muodostuvat tutkimusdatasta, tutkimusjulkaisuista, tutkimusmenetelmistä sekä niiden metatiedoista. Metatietojen ja objektien elinkaaren hallinnasta kertovan saatavuussuunnitelman pohjalta voidaan varmistaa aineistojen informaatioisällön muuttumattomuus säilytystoimenpiteissä. Tutkimusaineistojen hyödynnettävyyden kannalta pitkäaikaissaatavuuden turvaaminen tarkoittaakin erityisesti sisällön ymmärrettävyyden varmistamista ja sen vaatimien säilytysmenetelmien soveltamista. Tällöin voidaan taata, että tulevat käyttäjät pystyvät avaamaan aineistot, tulkitsemaan sisällöt ja hyödyntämään niitä kyseisenä aikakautena käytössä olevilla välineillä. Sisällön ymmärrettävyyden säilyttäminen asettaa selkeät vaatimukset aineistojen metatiedoille ja saatavuuden suunnittelulle, jotka tulee laatia ennen säilytyksen aloittamista. Ymmärrettävyyden säilyttämiselle perustan luo bittien säilyttäminen, joka suunnitelmallisesti tehtynä takaa vasta digitaalisen datan säilymisen.

## TOIMINTAMALLI

Tutkimuksen tuotosten pitkäaikaissaatavuuden turvaaminen on kaikkien tutkimusjärjestelmän toimijoiden yhteinen ponnistus tutkimuksen kestävä perustan turvaamiseen. Hyvien valintojen ja ratkaisujen tekemisen tueksi tarvitaan kansainvälinen tilannekuva sekä selvitykset toimintamallivaihtoehdoista tutkimusjulkaisujen ja tutkimusmenetelmien osalta. Määrittely- ja standardointityötä tarvitaan niin toimintamallin kuin pitkäaikaissäilytettävien sisältöjen osalta. Erityisen tärkeää on laatia yhdessä vaatimusmäärittelyt tarvittaville yhteisille palveluille.

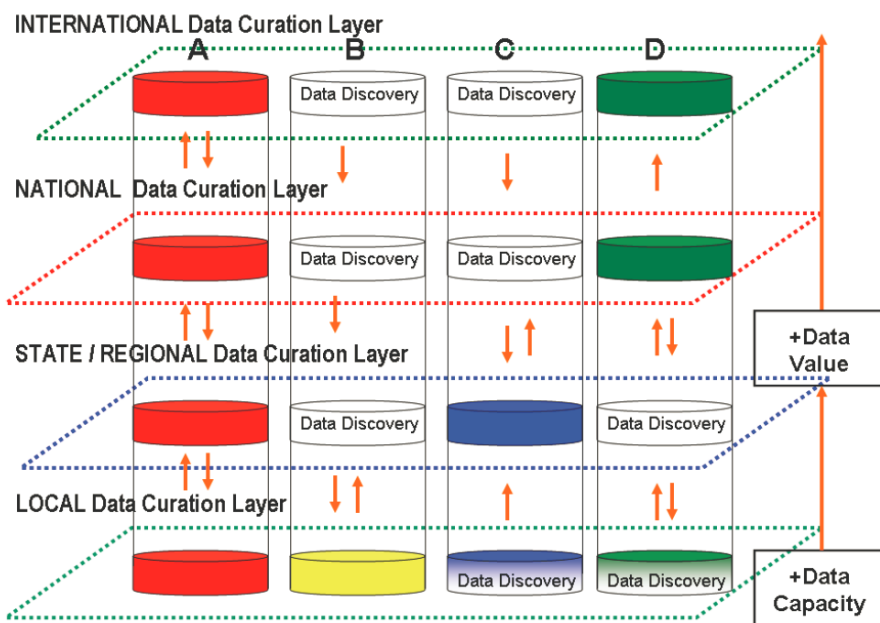
Pitkäaikaissaatavuutta tukevan kokonaisuuden tulee mahdollistaa ja tukea eri tieteenalojen ja toimintojen tarpeita, sekä olla muokattavissa oleva kerroksellinen, modulaarinen ja kustannustehokas ratkaisumalli, jossa hyödynnetään olemassa olevia komponentteja. Kokonaisuuden tulee mahdollistaa tutkimuksen tuotosten avoimuus ja jatkokäyttö primäärilähteistään.

Pitkäaikaissaatavuutta tukevassa toimintamallissa

- (1) Tutkimusmateriaalit ja tuotokset ovat aidosti kohdeyleisön käytettävissä,
- (2) Varaudutaan kustannustehokkaaseen ja läpinäkyvään ihmisten, organisaatioiden ja teknologioiden vaihtumiseen,
- (3) Varmistetaan, että
  - Bitit ovat tallessa ja avattavissa mielekkäästi;

- Metadatan kuvailutiedot ovat laadukkaat teknisesti ja riittävät ymmärrettävyyden varmistamiseen

Pitkäaikaissaatavuudessa on useita toiminnallisia kerroksia. Eräs taso on aineiston kohdeyleisön laajuus. Tutkimuksen kohdeyleisön pitkäaikaissaatavuuden yhteistyön mallia ovat hahmotelleet McDonald ja Waters (McDonald 2008). Ha erottavat neljä eri tapausta: kohdassa A aineistoihin on laaja kansainvälinen kiinnostus, kohdassa B aineistolla on arvoa ainoastaan paikalliselle organisaatiolle, kohdassa C kyseessä on suuri aineisto, jota ei pysty säilyttämään eikä kuratoimaan paikallisesti ja kohdassa D tarkastellaan laajaa aineistokokoelmaa, jota säilytetään jonkin aikaa paikallisesti mutta se siirtyy kansainvälisiin arkistoihin pitkällä aikavälillä.



Näitä erilaisia tasoja tukemaan tarvitaan yhteisiä toimintamalleja ja palveluita. Toimintamallin tulee kuvata miten tasolta toiselle siirryttäessä toimitaan. Kansainvälinen yhteentoimivuus tulee varmistaa sopivalla yhteistyöllä, sopimuksilla ja semanttisen tason tietomalleilla.

Myös oikeudellinen toimintaympäristö ja sen muutokset (mm. tekijänoikeuksien modernisointi Euroopan tasolla) vaikuttaa toimintamalliin.

Toimintamallin tulisi kuvata läpinäkyvästi eri tasojen vastuut, palvelun sisältö, kustannusten muodostuminen ja jakautuminen sekä etenemissuunnitelma. Toiminnan seurannan tulee olla helppoa ja ymmärrettävää niin organisaatio-, aineisto- kuin kokoelmakohtaisesti, samoin kootusti. Yhteisten palveluiden ja kansallisen tason palveluiden tulee mahdollistaa erilaiset liiketoimintamallit ja erilaiset exit-suunnitelmat niin organisaatio- kuin aineistokohtaisesti.

Toimintamallin periaatteiksi ehdotetaan:

- Eheys turvataan jo varhaisessa vaiheessa.** Syntyvien tutkimuksen tuotoksien saatavuus pyritään takaamaan eheyden turvaavalla yhteisellä palvelulla (ns. PAS-puskuri). Tuotoksia ei tässä vaiheessa arvoteta ja valikoida. PAS-puskurin säilytykseen laaditaan sopimus, jossa sovitaan vastuut.
- Automatisointiin pyritään.** Tutkimuksen tuotoksien käytön ja säilytystarpeen arviointi pyritään tekemään mahdollisimman automaattiseksi (viittauksien ja lataamisten ja katselukertojen tilastointi, uusien versioiden synty), eikä henkilötyöstä koostuvaksi. Säilytys sopimus laaditaan.

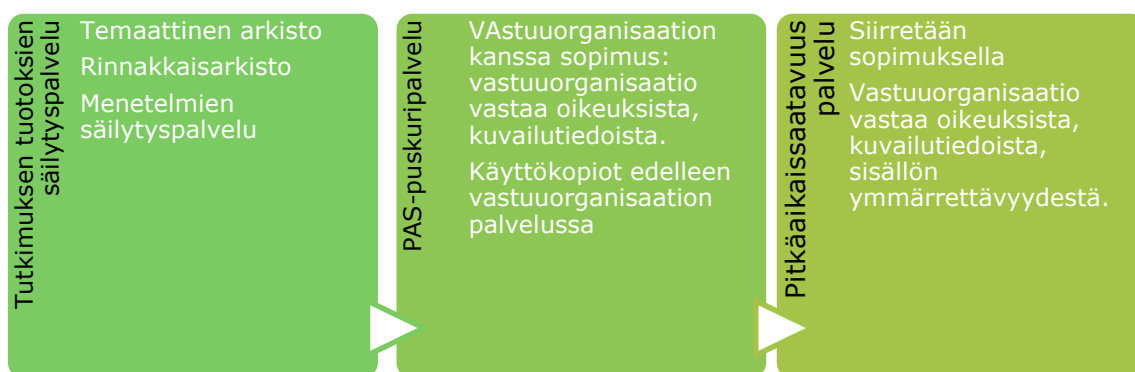
3. **Ainutlaatuiset aineistot turvataan.** Uniikit aineistot voidaan siirtää pitkäaikaissaatavuuden piiriin heti.
4. **Harkittu siirto tutkimus-PAS –palveluun.** Kymmenen vuoden kuluttua suoritetaan arvio, sekä siirretään käytetyt tuotokset pitkäaikaissaatavuuden takaavaan palveluun, joka turvaa eheyden lisäksi ymmärrettävyyden. Tällöin ne poistetaan ns. PAS-puskuripalvelusta. Säilytys sopimus laaditaan.
5. **Vastuut nimetään.** Säilytys sopimuksia laadittaessa määritetään vastuuorganisaatio. Se voi olla aineistot tuottanut organisaatio, tai jokin temaattinen keskittymä. Vastuuorganisaatiolla tulee olla tuotoksiin sellaiset oikeudet, että se voi solmia sopimuksen. Vastuuorganisaatio vastaa aineistojen käyttöoikeustietojen oikeellisuudesta. Vastuuorganisaatio vastaa kuvailutietojen laadusta. Kuvailutiedot ovat kaikkien saatavilla, paitsi siltä osin kun ne sisältävät ei-julkista tietoa. Vastuuorganisaatioita tuetaan PAS-työssä monin tavoin.
6. **Kustannusten kattamisesta sovitaan.** Kustannusten kattamisesta pyritään sopimaan tutkimuksen toimijoiden kesken, niin etteivät kustannukset kohdistu yksittäisiin organisaatioihin.

Näiden periaatteiden vaikutus hallinnolliseen suunnitteluun, tietotekniseen kehittämiseen ja palveluun ovat seuraavat:

- **Hallinnollinen suunnittelu:** Suunnittelussa keskitytään eheysuhan alla olevien aineistojen tunnistamiseen, ainutlaatuisten aineistojen tunnistamiseen, vastuutahojen tunnistamiseen sekä kustannuksista sopimiseen.
- **Tekninen kehittäminen:** Kehittämisessä painotetaan prosessivaiheiden täsmennystä ja niiden automatisointia.
- **Palvelun kehitys:** Keskitytään hierarkkisen yhteentoimivan palvelun edistämiseen palvelukokonaisuudessa.

Erilaisia toimijoita tukeva Tutkimus-PASin palvelutoiminta olisi kolmiportainen:

1. Paikallinen tai temaattinen säilytyspalvelu säilyttää syntyneitä tutkimusaineistoja
2. Säilytyspalvelusta siirretään tarvittaessa aineistoja ns. PAS-puskuripalveluun
3. Varsinaiseen pitkäaikaissaatavuuspalveluun aineistoja siirretään, kun niiden säilytyskelpoisuus – ja tarve on varmistettu.
- 4.



Näihin eri tasoihin liittyviä palveluita kuvataan tarkemmin kohdassa "Toimintamallin arkkitehtuuri".

## TOIMIJIAT

Pilottien kokemuksen sekä muiden selvitysten pohjalta Tutkimuksen PAS palvelun toimintamalli eroaa KDK-PAS palvelun toimintamallista ja palvelun piirissä oletetaan jatkossa olevan useita eri toimijoita:

**(1) Organisaatiot, joiden vastuulla on paikallisia arkistoja ja tutkimuksen tuotoksia**

- a. vastaavat aineistojen, julkaisujen ja menetelmien ja niihin liittyvien metatietojen säilytyksestä
- b. vastaavat metatietojen toimittamisesta kokoaviin palveluihin

**(2) Tutkimusaineistoja tuottavat yksiköt, kuten tutkimusryhmät, jotka tuottavat tutkimusaineistoja satunnaisesti**

- a. vastaavat tuotosten ja niihin liittyvien metatietojen siirtämisestä paikallisiin, kansallisiin tai kansainvälisiin palveluihin

**(3) Tutkimusaineistojen ymmärrettävyyden takaamisesta vastuussa olevat toimijat**

- a. vastaavat aineistojen sisällön ymmärrettävyydestä paikallisissa tai kansallisissa palveluissa, erikoissopimuksesta myös kansainvälisissä palveluissa

**(4) Kansallisen tason yhteentoimivuudesta ja palveluista vastaavat toimijat**

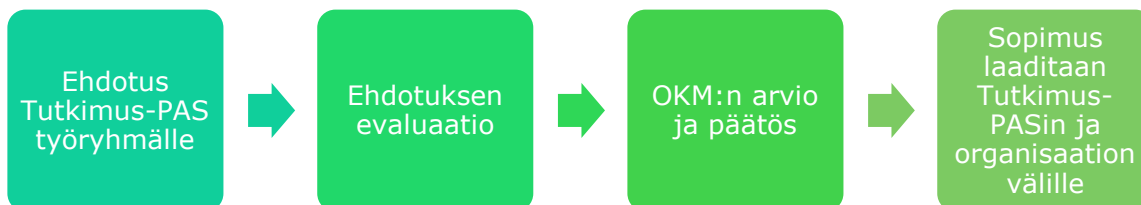
- a. vastaavat metatietojen yhteentoimivuudesta
- b. vastaavat metatietotojen kokoamisesta ja yhteisistä näkymistä
- c. vastaavat pitkäaikaissaatavuuteen liittyvistä säilytys- ja saatavuuspalveluista
- d. vastaavat kansainvälisen tason yhteentoimivuudesta ja tietojen vaihdosta

Lähtökohtaisesti aineistoihin liittyvät oikeudet eivät muutu pitkäaikaissaatavuuspalvelun eri tasoilla. Kun tutkimusryhmien kokoonpano muuttuu tai tutkimusprojekti päättyy, pitkäaikaissaatavuudeltaan turvatus tutkimusaineiston ymmärrettävyyden säilyminen ei välttämättä ole kenenkään vastuulla, jollei tähän tehtävään ole nimetty vastuutaho, jolle myös oikeudet siirtyisivät. Oletettavaa onkin että syntyy tutkimuksen tuotosten ymmärrettävyyden ylläpidon keskittymiä (kuten tutkimusinfrastruktuurit, tutkimuslaitokset), jotka ottavat vastuulleen päättyvien projektien tuottamien tutkimusaineistojen ymmärrettävyyden säilyttämisen sekä muut aineiston elinkaaren liittyvät kysymykset kuten oikeudelliset ja tietosuojan liittyvät kysymykset.

*Vuosien 2016-2017 aikana on tärkeä sopia näiden tehtävien vastuista, esimerkiksi muodostamalla temaattisia keskittymiä. Aloitetaan myös keskustelut kansainvälisten toimijoiden kanssa työnjaosta.*

## SOPIMUKSET

Sopimuksien laatiminen lähtee liikkeelle ehdotuksesta aineistokokonaisuuden saattamisesta Pitkäaikaissaatavuuspalveluun. Ehdotus tehdään Tutkimus-PAS -työryhmälle (OKM:n osoittaman vastuuhenkilön kautta). OKM:n myönteisen arvion jälkeen laaditaan sopimus.



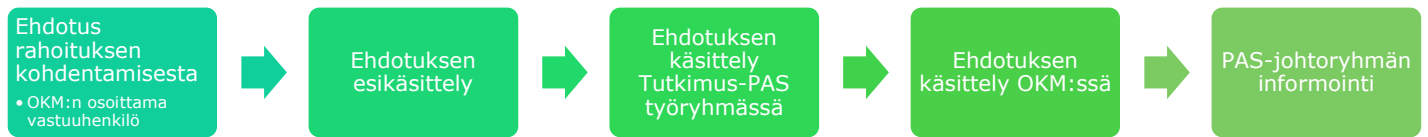
Mikäli eteneminen vaatii resursoinnin kohdentamista, toimitaan seuraavan kohdan mukaisesti. Jos mahdollista, resursointipyyntö tulisi esittää sopimuksen yhteydessä.

*Vuoden 2016 aikana laaditaan alustava sopimusmalli.*



## RESURSOINTIPÄÄTÖKSET

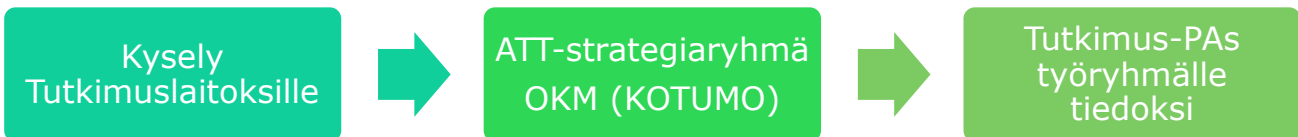
Resursoinnin kohdentamisesta sopiminen etenee seuraavasti:



Resursointi voi olla esimerkiksi pilotoinnin tukemista organisaatiossa tai säilytys oikeus PAS-puskuripalvelussa, tai käyttökopioiden jakelupalvelussa.

## KUSTANNUSTEN HALLINTAMALLI

Tavoitteena on, että pitkäaikaissaatavuuspalvelu on tutkijoille helppo ja kannustava tapa toimia, ja aineistojen elinkaaresta huolehtiminen on luonteva osa tutkimuksen käytäntöjä. Pitkäaikaissaatavuuden takaavassa toimintamallissa pyritään löytämään kansallisten, yhteisten palveluiden kattamiseen hallintamalli, jonka avulla kustannukset voidaan maksaa sektoreittain kootusti, eivätkä ne koituisi yksittäisten organisaatioiden tai yksiköiden vastuulle.

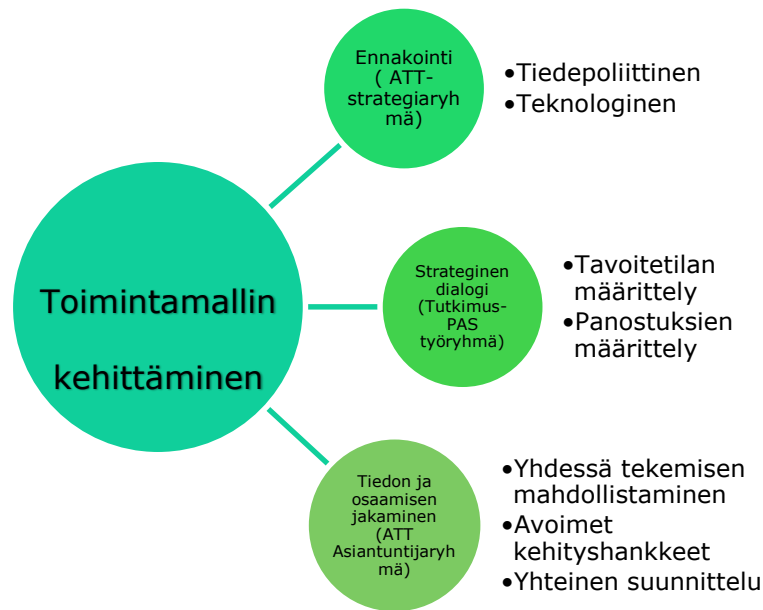


*Vuoden 2016 aikana laaditaan alustava kustannusten jakomalli.*

## TOIMINTAMALLIN KEHITTÄMINEN

Pitkäaikaissaatavuuden turvaaminen tarkoittaa jatkuvaa uudistumista kentän ja toimijoiden muutosten mukana. Toimintamallin tulee olla vahva yhteisöllisen osaamisen kiteytyminen ja sen tulee mahdollistaa sektoraalet raja-aidat ja tieteenalat ylittävä ajattelu ja yhteistyö. Tämä tarkoittaa siirtymistä resurssien ja palveluiden omimisesta kohti kokonaisuuden hyvää toimintaa ja kaikkia tukevaa mallia. Muutokset ovat nopeimpia tutkimusprosessin ytimessä, ja ne implementoidaan hitaammin pitkäaikaissaatavuuspalveluihin. Muutoksista täytyy kuitenkin olla PAS-kontekstissa tietoisia, koska ne voivat vaikuttaa jäljitettävyyteen.

Toimintamallin kehittämiseen kuuluvia osiota on kuvattu alla. Toimintamallin kehittämisen tulee olla jatkuvaa toimintaa.



Laaja yhteistyö eri toimijoiden ja verkostojen kanssa takaa toiminnan kehittymisen tutkimusta yhä paremmin tukevaan muotoon. Tavoitteellisessa kehittämisessä pyritään ketteryyteen ja osallistamiseen.

Oleellinen osa toimintamallia on toimintaa liittyvien uhkien tunnistaminen ja riskien hallinta. Riskiä määriteltäessä tulisi tarkastella epätoivotun seuraamuksen haitallisuutta ja todennäköisyyttä. Riski toteutuu yleensä vaaralle altistumisen seurauksena ja sen hyväksyttävyys riippuu monista tekijöistä. Pitkäaikaissaatavuuden turvaamisessa toimintamallin riskianalyysi täytyy tehdä koko toimintaketjun osalta.

*Vuoden 2016 aikana on hahmoteltava toimintamallin kehittämisen jatko ATT-hankkeen päättymisen jälkeen 2017.*

*Vuoden 2016 aikana on tehtävä alustava riskianalyysi tutkimuksen tuotosten pitkäaikaissaatavuuden turvaamisen osalta.*

## TUNNISTETTUJA TUTKIMUSAINEISTOJA

Jotta laadulliset hyödyt saadaan realisoitua, Tutkimuksen tuotosten pitkäaikaissaatavuuden turvaamisessa täytyisi tunnistaa tärkeimmät tutkimuksen tietovarannot. Syksyllä 2015 pyydettiin ATT-hankkeen ohjausryhmiltä ehdotuksia strategisesti tärkeistä tutkimusaineistoista. Alla kooste saaduista aineistotiedoista, joilla on pitkäaikaissaatavuuden ulottuvuus. Tärkeitä aineistoja on toki noussut esiin jo aiemmin piloteissa, sekä Tutkimusinfrastruktuurien tiekartalla.

FSD, Suomi-Monitori
Turun yliopisto 1, IT-barometridata
Turun yliopisto 2, EVEX Biomedical data
Lahden AMK, FUAs hyvinvointi (PAS-pilotti)

LYNET 1, Mikroöbilääkeresistenssin seurantatutkimus
LYNET 2, Lypsykarjojen tautiseurantatutkimus
Jyväskylän yliopisto 1, viittomakielen korpus (PAS-pilotti)
THL 1, Aikuisväestön ja Alueellinen terveys- ja hyvinvointitutkimus
THL 2, Perusterveydenhuollon avohoidon hoitoilmoitus
THL 3, Implanttirekisteri
Helsingin yliopisto 1, SuALT The Finnish Archaeological Finds recording Database
Helsingin yliopisto 2, FIRE - luotusaineistot
Helsingin yliopisto 4, Sanomalehtiaineistoa tutkijoille
Helsingin yliopisto 5, Digitaalinen murrearkisto
Helsingin yliopisto 6, von Wright and Wittgenstein archives
Helsingin yliopisto 7, Kriminologian ja oikeuspolitiikan instituutin aineistot

*Vuoden 2016 aikana tulee tutkimusaineistoja tunnistaa lisää mm. pilotointien kautta.*

## TOIMINTAMALLIN ARKKITEHTUURI

ATT-kokonaisuuden tietojärjestelmäpalvelut ja sähköinen palveluympäristö jäsennetään tavoitetilassa soveltaen SOA-periaatteita tarkoituksenmukaisessa laajuudessa.

Palvelukeskeinen arkkitehtuuri (Service-Oriented Architecture, SOA) antaa mahdollisuuksia sekä eristää tietojärjestelmäpalveluita erillisiksi toimiviksi osiksi että yhdistää palveluita tarvittaessa. Järjestelmien toiminnallisuus (liiketoimintalogiikka) rajataan pääsääntöisesti ohjelmallisen rajapinnan taakse – vastakohtana palvelut, joita voi käyttää vain loppukäyttäjän käyttöliittymän kautta. Tällöin on mahdollista hyödyntää kyseistä tietojärjestelmäpalvelua ohjelmallisesti yhteiskäyttöisiä komponentteja muissa palveluissa ja toiminnoissa. Avoimella dokumentoidulla rajapinnalla tarkoitetaan rajapintaa, jonka teknologiana on käytetty jotain yleisesti tiedossa olevaa ja dokumentoitua protokollaa ja tiedostomuotoa. SOA-palveluissa olennaista on myös rajapinnan kautta siirrettä-vien tietojen semanttinen yhteentoimivuus.

ATT-viitearkkitehtuurissa keskeisimmät tietojärjestelmäpalvelut perustuvat tavoitetilassa toisiinsa löyhästi kytketyistä tietojärjestelmäpalveluista tai tietojärjestelmäpalvelujen taakse ”piilotettuja” järjestelmiä.

Tärkeimmät ATT-tietojärjestelmäpalveluiden suunnittelussa ja toteutuksessa noudatettavat SOA-ratkaisuperiaatteet ovat:

- Uudelleenkäytettävyys
- Systemaattisuus

- Palvelevuus toiminnan näkökulmasta
- Tekninen ja semanttinen yhteentoimivuus
- Joustavuus ja laajennettavuus

---

## PROSESSIEN KOKOAMINEN

Avoimen tieteen ja tutkimuksen viitearkkitehtuuri perustuu siis osittain hajautettuun arkkitehtuuriin, jossa osa tiedoista ja aineistoista säilytetään hajautetusti tutkimuslaitosten omissa tietovarannoissa ja palveluissa. Näistä poimitaan metatiedot keskitettyyn palveluun, josta ne ovat kokonaisuudessaan helposti haettavissa ja latauspalvelun kautta ladattavissa:

Osa tiedoista ja palveluista kannattaa kuitenkin keskittää kansallisiin palveluihin. Näitä ovat tyypillisesti sellaiset palvelut, joihin korkeakouluilla ja tutkimuslaitoksilla ei vielä ole laajasti valmiita palveluja tai tutkimusaineistoja.

Kun pitkäaikaissaatavuuden turvaamiseen liittyviä prosesseja kootaan, tulee pyrkiä:

- **Läpinäkyvyyteen toimijoille.** Läpinäkyvyys on keino toimijoiden luottamukseen ja sitoutumiseen.
- **Monimuotoisuuteen ja hajautettuun arkkitehtuuriin.** Ihanteellisen pitkäaikaissaatavuuden kokonaisuuden tulisi mahdollistaa monimuotoisuus kaikilla tasoilla. Tutkimus-PAS – kokonaisuudessa tätä tuetaan ATT-viitearkkitehtuurilla, joka rakentuu hajautettuun ratkaisuun ja monitasoiseen kehitykseen SOA-ajattelun pohjalta.
- **Kustannustehokkuuteen ja talouden hyvään hallintaan.** Kustannuksien hallintaa tulee tehdä kaikilla tasoilla mm. prosesseja digitalisoimalla ja säilytettävien tutkimusmateriaalien hyödyntämistä tehostamalla sekä oikein kohdistetuilla resursointipäätöksillä (mm. hankinnat).
- **Erialaisten liiketoimintamallien tukemiseen** (esim. kv-yhteistyössä). Osa palveluista tai prosesseista voidaan toteuttaa kansainvälisten toimijoiden kanssa tai PPP-yhteistyössä. Liiketoimintamallin tulee mahdollistaa parhaan ratkaisun hyödyntäminen.
- **Erialaisten exit-ratkaisujen tukemiseen.** Kansalliselta alueelliselle, kansalliselle ja kansainväliselle tasolle siirryttäessä materiaalien siirto palvelusta toiseen tulee olla helppoa.
- **Seurattavuuteen ja jäljitettävyyteen** organisaatio-, aineisto- ja kokoelmakohtaisesti. Nämä mahdollistavat laadun seurannan ja validoinnin.

*Vuoden 2016 aikana tulee pyrkimykset ja toimintamallin periaatteet sisällyttää vaatimusmäärittelyihin.*

---

## PROSESSIEN YHTEENTOIMIVUUS

Päätiedot/ydintiedot tunnustetaan kytkemällä käsitelmä avoimen tieteen ja tutkimuksen prosessiin. Ne ovat loogisia kokonaisuuksia, joihin niitä koskevat ominaisuudet (attribuutit) kiinnitetään. Päätiedot ovat niitä keskeisimpiä prosessissa eri toimijoiden, tietojärjestelmäpalvelujen ja tutkimusaineistojen välillä liikkuvia tietoja, joiden yhteentoimivuus on erityisen tärkeää. Näkökulma on sekä prosessin yhteentoimivuutta varmistava että tiedonhallinnallinen. Yhteentoimivuuden kehittäminen kannattaa aloittaa päätiedoista.

Kokonaisuuden tulee mahdollistaa tutkimuksen materiaalien tuotosten avoimuus ja jatkokäyttö primäärilähteistään. Kokonaisuuteen kuuluvat siis:

- tutkimusjulkaisut
- tutkimusdata
- tutkimusmenetelmät (mukaan lukien tutkimussuunnitelmat)
- julkaisuja, dataa ja menetelmiä koskeva metadata, laatutieto
- julkaisujen, datan ja menetelmien väliset viittaukset ja linkit
- käyttöoikeustiedot (esimerkiksi lisenssit).

Pitkäaikaissaatavuuden takaava toimintamalli rakentuu semanttiselle yhteentoimivuudelle. Tunnisteiden jakaminen avointen rajapintojen kautta vahvistaa tätä. Tällöin

- Tehokas tiedonhallinta edellyttää tietosisältöjen yhteismitallisuutta. Tämä tapahtuu viittaamalla eri *järjestelmien välillä* yhteisiin tunnisteisiin
  - Tunnisteet ovat edellytys löydettävyydelle ja yhteentoimivuudelle
  - Tunnisteita on erilaisia eri käyttötapauksiin
  - Laadukkaat tunnisteet ovat pysyviä
  - Eri järjestelmissä olevat tunnisteet samoille objekteille on syytä linkittää
  - Keskeisen tietovarannon ylläpitäjän tulisi tarjota pysyviä tunnisteita omille objekteilleen avoimesti ja koneluettavasti
  - Ainoa hyväksyttävä rajoitus tunnisteiden avoimuudelle on juridinen este (esim. tietosuoja)
  - Esim. yrityksen tunniste voisi olla uri-alkuinen Y-tunnus. Rajapinnan kautta kaikki järjestelmät voisivat hakea yksiselitteisesti yrityksen perustiedot
- Linkitetty data edellyttää, että tutkimusaineistojen omistajat vastaavat tunnisteiden pysyvyydestä, ajantasaisuudesta ja saatavuudesta

---

## TOIMINTAMALLIN PROSESSIT

Toimintamalli liittyy korkeakoulujen ja tutkimuslaitosten normaaliin toiminnan kokonaisuuteen. Näitä prosesseja on kuvattu tavoitetasolla ATT-hankkeen viitearkkitehtuurissa sekä palvelukartassa. Pitkäaikaissaatavuuden turvaamisen toimintamalliin kuuluvat alustavasti seuraavat prosessit:

### - **Aineistojen hallinta**

**Kattaa:** Aineistojen paikallinen ja alueellinen hallinta, liittyen mm. laatuprosesseihin ja yhteentoimivuustyöhön

- Aineistojenhallinnan suunnittelu. Palvelukehitys: Tuuli (ATT ja HY).
- Esikäsittely, analyysi ja prosessointi.
- Aineistojen kokoaminen ja suostumusten hallinta.
- Aineistojen paikallinen ja alueellinen säilytys ja arkistointi.
- Yhteentoimivuus (rajapintapalvelu).
- Aineistojen käytön seuranta ja raportointi.

### - **Aineistojen metatietojen hallinta**

Kattaa: Aineistojen kuvailu tunnisteita, sanastoja ja ontologioita hyödyntäen, sekä syntyneiden metatietojen jakaminen.

- Kuvailu.
- Palvelukehitys
- Tunnisteiden, luokitusten ja ontologioiden hallinta
- Metatietojen jakaminen
- Hakupalvelu
- Tarvittavien näkymien koostaminen hakuihin paikallisella, alueellisella tai kansallisella tasolla
- Käytön seuranta ja raportointi, viitteiden hallinta

#### - **Säilytyksen suunnittelu**

Kattaa: Elinkaaren suunnittelu ja siihen tarvittavien sopimusten teko ja resurssien hallinta.

- Aineiston hallinnan suunnittelu
- Säilytyksestä sopiminen
- Rahoituksen ja resurssien seuranta

#### - **Vastaanotto**

Kattaa: Pitkäaikaissäilytyspalveluun siirtyvien aineistojen valmistelu ja siirto.

- Kuvailutietojen kerääminen
- Automatisoitu aineiston paketointi ja vastaanotto

#### - **Pitkäaikaissäilytys**

Kattaa: Aineistojen käytettävyyden takaavat teknologiset toimenpiteet.

- Eheyden tarkastaminen
- Virkistys ja replikointi
- Tarvittavat migraatiot yhdessä vastuutahojen kanssa
- Teknologiaseuranta

#### - **Käyttö:**

Kattaa: Aineistojen saattaminen käyttökelpoisessa muodossa uudelleenkäyttöön.

- Saatavuuspalvelu
- Käyttövaltuuspalvelu
- Käyttökopioiden hallinta
- Lisensointimallit
- Tutkimuseettiset ohjeet ja tutkimuslupien hallinta
- Käytön seuranta

#### - **Hallinnointi**

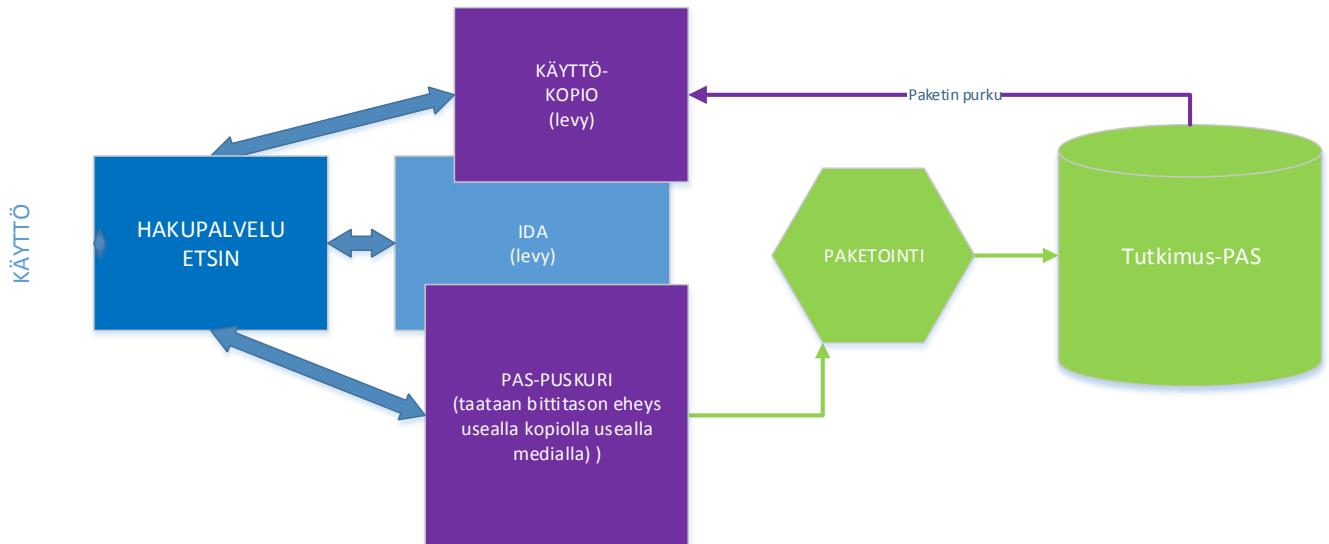
Kattaa: Kokonaisuuden tavoitteiden mukaisen toiminnan varmistaminen ja seuranta.

- Yhteentoimivuus
- Raportointi
- Sopimuksien toteutumisen seuranta
- Arviointi

*Vuoden 2016 prosesseja ja niiden kytköksiä tulee täsmentää ATT-tavoitearkkitehtuurin ja palvelukartan pohjalta.*

## ALUSTAVA YHTEISEN PALVELUARKKITEHTUURIN MALLI

Nykyisiä yhteisiä ATT-hankkeen palveluita tulee kehittää Pitkäaikaissaatavuuden turvaamisen mahdollistavaan malliin. Alkuvaiheessa tämä tarkoittaa IDA-palvelun kehittämistä toisaalta PAS-puskuriksi ja toisaalta käyttökopioiden hallintaan. Etsin-hakupalvelua kehitetään toimimaan kokonaisuuteen liittyvänä kattavana hakupalveluna. Puuttuvien osien tunnistaminen jatkuu ATT-arkkitehtuurityössä ja palvelutyöryhmän työssä vuonna 2016.



*Vuoden 2016 aikana tulee laatia roadmap nykyisten ATT-palveluiden kehitykselle.*

## YHTEENVETO

Tutkimuksen tuotosten pitkäaikaissaatavuuden turvaamisen toimintamallilla tarkoitetaan suunnitelmaa seuraavien 5-10 vuoden aikana tapahtuvasta uudistamistyöstä. Toimintamallin periaatteiden pohjalta suunnitellaan pitkäaikaissaatavuuden prosessit ja palvelut. Toimintamalli kytkeytyy korkeakoulujen ja tutkimuslaitosten sisäiseen toiminnan kokonaisuuteen, mutta myös kansainvälisiin tutkimusinfrastruktuureihin. Erityisesti yhteisiä kansallisia palveluita tulee kehittää näiden kytköksiin kautta. Vuoden 2016 aikana luodaan ATT-työssä pohja toiminnalle. Tähän kuuluvat seuraavat asiat:

- Toimintamalliin liittyvistä vastuista sopiminen, esimerkiksi muodostamalla temaattisia keskittymiä.
- Aloitetaan myös keskustelut kansainvälisten toimijoiden kanssa työnjaosta.
- Alustavan sopimusmallin laatiminen.
- Alustavan kustannusten jakomallin laatiminen.
- Sovitaan toimintamallin kehittämisen jatko ATT-hankkeen päättymisen jälkeen 2017.
- Tehdään alustava riskianalyysi tutkimuksen tuotosten pitkäaikaissaatavuuden turvaamisen osalta.
- Pyrkimykset ja toimintamallin periaatteet sisällytetään vaatimusmäärittelyihin.
- Tutkimuksen tuotosten saatavuuden turvaavia prosesseja ja niiden kytköksiä tulee täsmentää ATT-tavoitearkkitehtuurin ja palvelukartan pohjalta.
- Laaditaan roadmap nykyisten ATT-palveluiden kehitykselle
- Tärkeitä tutkimusaineistoja tulee tunnistaa mm. pilotointien kautta.

Viitteet:

Robert H. McDonald and Tyler O. Walters, 2008. Sustainability Models for Digital Preservation Federations. Chronopolis. (This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 2.5 License <<http://creativecommons.org/licenses/by-nc-sa/2.5/>>.)