

Pimeän arkiston toteutusvaihtoehtoja Theseukselle

26.4.2016

Samu Viita (samu.viita@helsinki.fi)

Kirjastoverkkopalvelut

Suunnittelun taustaa ja lähtökohdat

- Kansalliskirjastoa pyydetty laatimaan ehdotuksia pimeän arkiston toteuttamiseksi
- Muna-kana -ongelma:
 - Vaikea suunnitella, kun vaatimukset järjestelmälle puuttuvat
 - Vaikea laatia vaatimusmäärittely, jos ei tiedetä mitä halutaan, mikä mahdollista ja tulevatko kaikki mukaan
- Tämän vuoksi päädyttiin esittelemään kolme vaihtoehtoista mallia keskustelun avaamiseksi ja vaatimusmäärittelyn tueksi
- Lopullinen arkkitehtuuri voi olla joku näiden yhdistelmä tai jotain ihan muuta

Suunnittelun periaatteita ja reunaehdoja

- Käytettävä olemassa olevaa tekniikkaa ja toimintatapoja
Toteutusresurssit rajalliset
Nykyisillä taidoilla tulisi pärjätä myös pimeän arkiston kanssa kirjastoissa
Suunnittelua ei voida siis aloittaa puhtaalta pöydältä ->
Pitäydytään Dspace-järjestelmässä
- Kaikki vai osa mukana? Tämä vaikuttaa suunnitteluun
- Käytettävyys vs. tehokkuus (levynkäyttö, julkaisunopeus)
- Seuraavalla kolmella mallilla pyritään näyttämään erilaisia toteutuskelpoisia vaihtoehtoja.

Malli 1

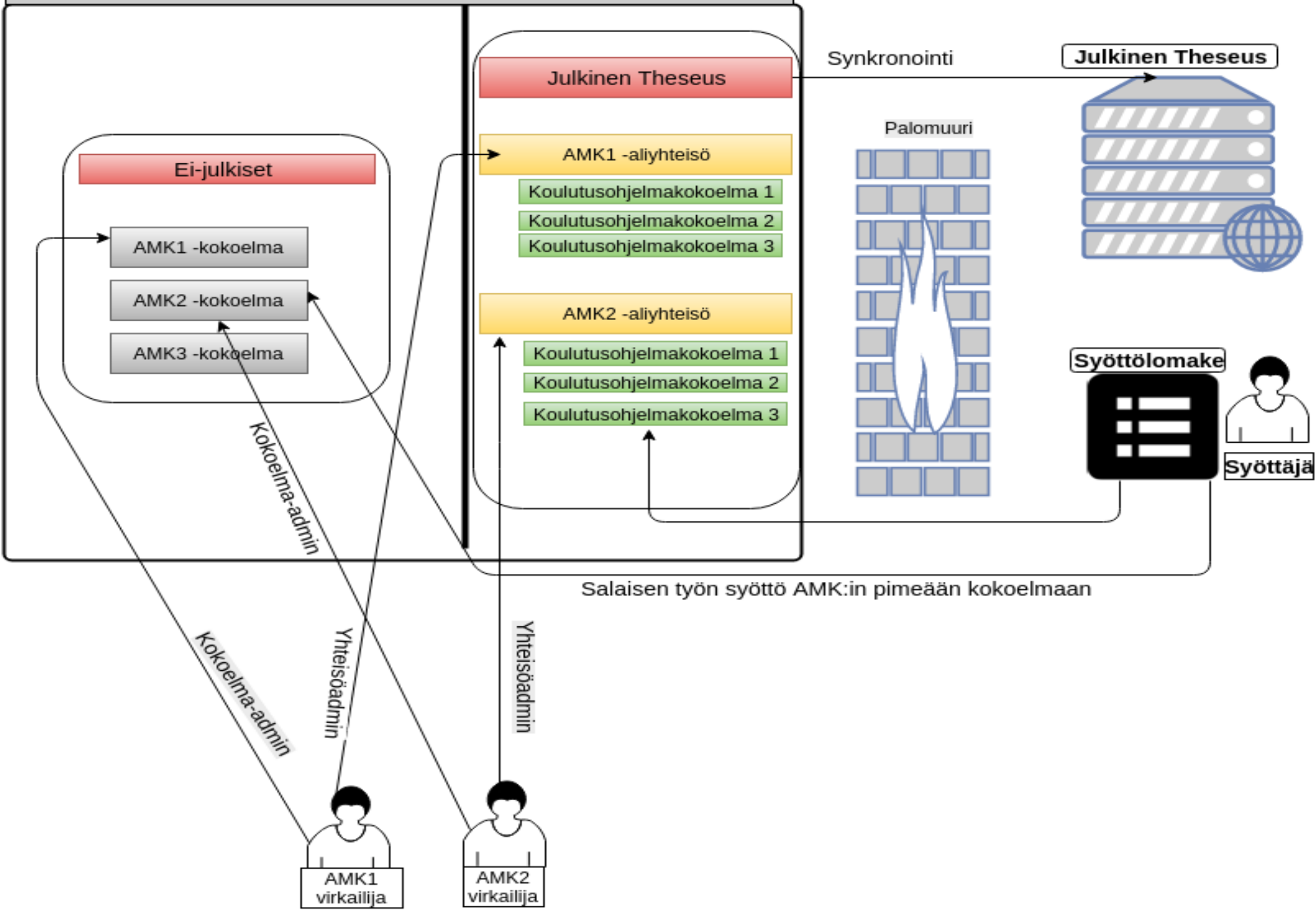
- Kokonaisuus koostuu kahdesta Dspace-instanssista (kuten kaikkien muidenkin esiteltävien mallien)
- Kaikki tietueet, ei-julkiset ja julkiset, sijaitsevat pimeässä arkistossa, palomuurin takana
- Tämän lisäksi julkinen Theseus-instanssi, johon pimeän arkiston julkiset tietueet siirtyvät automaattisesti
- Virkailijat operoivat pelkästään pimeän arkiston puolella
- Ei reaaliaikaista julkaisua: syötön ja julkiseen Theseukseen siirtymisen välillä on viivettä
- Malli **mahdollinen vain, jos kaikki tulevat mukaan**, koska operointi tapahtuu vain pimeässä arkistossa

Malli 1

- Kaikilla AMK:eilla **yksi** kokoelma, johon ei-julkiset tietueet siirretään. Kokoelma toimii varsinaisena pimeänä arkistona AMK:lle
- Siirto pimeään ja julkisen välillä onnistuu samalla käyttöliittymällä, mikä on nytkin Theseuksessa
- Julkiset tietueet tallentuvat levyille kahteen kertaan: pimeään arkiston julkiseen osioon ja varsinaiseen julkiseen instanssiin (Theseukseen)
- Siirto pimeään arkiston julkisesta kokoelmasta Theseukseen automaattista, nopean poiston ja päivityksen suhteen automatiikan selvittely kesken
- Opinnäytesyöttölomake kytketty pimeään arkiston puolelle, workflowtoiminto sama kuin ennen

Theseuksen pimeä arkisto

Yksi Dspace instanssi, kaksi pääyhteisöä



Malli 1 plussat ja miinukset

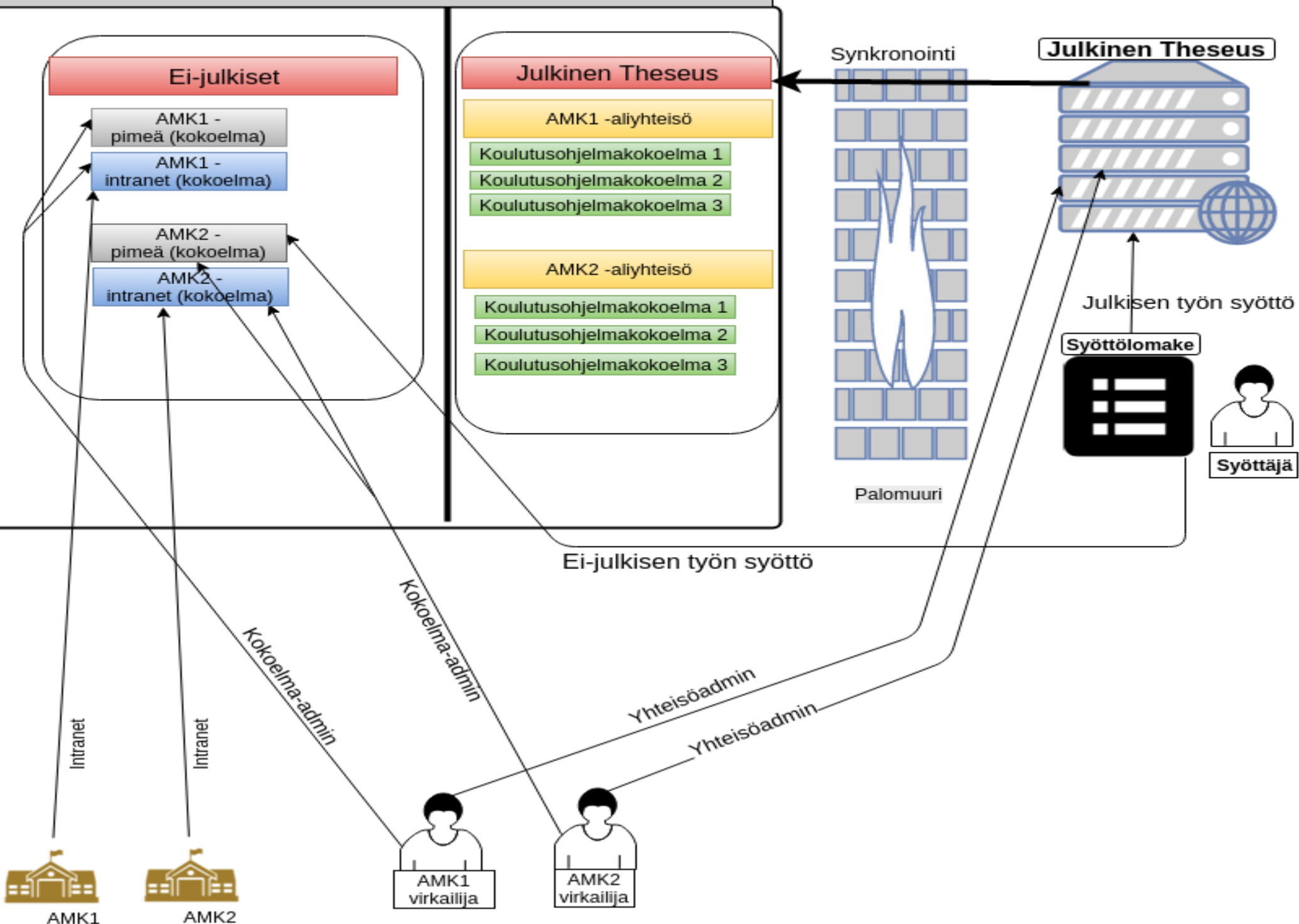
- + Yksi tunnus riittää virkailijoille
- + Selkeä ja turvallinen kokonaisuus
- + Pimeän ja julkisen välillä tehtävät siirrot helppokäyttöistä
- + Kaikki tietueet löytyvät yhdestä järjestelmästä / instanssista (Tietueet haettavissa, siirrettävissä ja varmistettavissa kattavasti)
- Mahdollinen vain jos kaikki AMK:t mukana
- Julkiset tiedostot tallennetaan kahteen kertaan
- Julkaisussa viivettä julkiselle puolelle siirron kanssa
- Mahdollista, että julkiselta puolelta poistojen osalta poisto tehtävä molemmissa järjestelmissä. Tämä asia vielä selvittelyssä

Malli 2

- Kaksi instanssia, kuten mallissa 1
- Kaikkien ei tarvitse olla mukana
- Theseuksen käyttö jatkuu samana, pimeä arkisto vain ekstra
- Syöttölomake kytketty julkiseen Theseukseen
- Synkronointisuunta eri kuin ykkösmallissa: julkisesta pimeään
- Pimeässä arkistossa kaikki tietueet, (julkiset ja ei-julkiset) niiden osalta, jotka mukana
- Mallissa 2 esitellään myös intranet-vaihtoehto. Tämä onnistuisi myös mallissa 1
- Intranet tuo lisää ylläpitoa IP-rajauksien osalta

Theseuksen pimeä arkisto

Yksi Dspace instanssi, kaksi pääyhteisöä



Malli 2 plussat ja miinukset

- + Kaikkien ei tarvitse olla mukana, Theseuksen käyttö tällöin mahdollista entiseen tapaan
- + Pimeän arkiston hankkineiden osalta kaikki tietueet löytyvät pimeästä arkistosta
- Virkailija tarvitsee tunnuksen molempiin järjestelmiin (pimeän arkisto käyttäjien osalta)
- Syöttölomake vaatii muutoksia, lisätoimintona mahdollisuus tallettaa pimeään syöttövaiheessa (pimeän arkisto käyttäjien osalta)
- Ei-julkinen työ voi lipsahtaa julkiselle puolelle, koska ohjaus tapahtuu syöttövaiheessa opiskelijan toimesta
- Tietueet tallennetaan kahteen kertaan

Malli 3

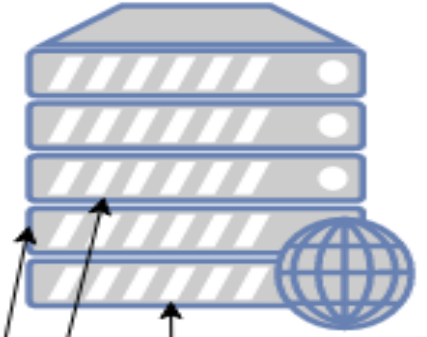
- Kuten malli 2, mutta pimeä arkisto ei sisällä julkisia töitä
- Syöttö julkiselle puolelle
- Theseusta käytetään kuten ennenkin, mutta pimeä arkisto on mahdollista ottaa käyttöön niin halutessa
- Samat plussat ja miinukset kuin mallissa 2, paitsi että tietoja ei tallenneta kahteen kertaan julkisten osalta. Tämä sekä plussa, että miinus:
 - + Levytilaa säästyy, kevyt ratkaisu
 - Pimeä arkisto ei sisällä kaikkia tietueita, vaan ne täytyy haalia kokoon esim. luovutuksia varten kahdesta järjestelmästä.

Theseuksen pimeä arkisto

Ei-julkiset

- AMK1 -pimeä (kokoelma)
- AMK1 -intranet (kokoelma)
- AMK2 -pimeä (kokoelma)
- AMK2 -intranet

Julkinen Theseus



Julkisen työn syöttö



Syöttölomake



Syöttäjä



Palomuuuri

Salaisen työn syöttö

Intranet

Intranet

Kokoelma-admin

Kokoelma-admin

Yhteisöadmin

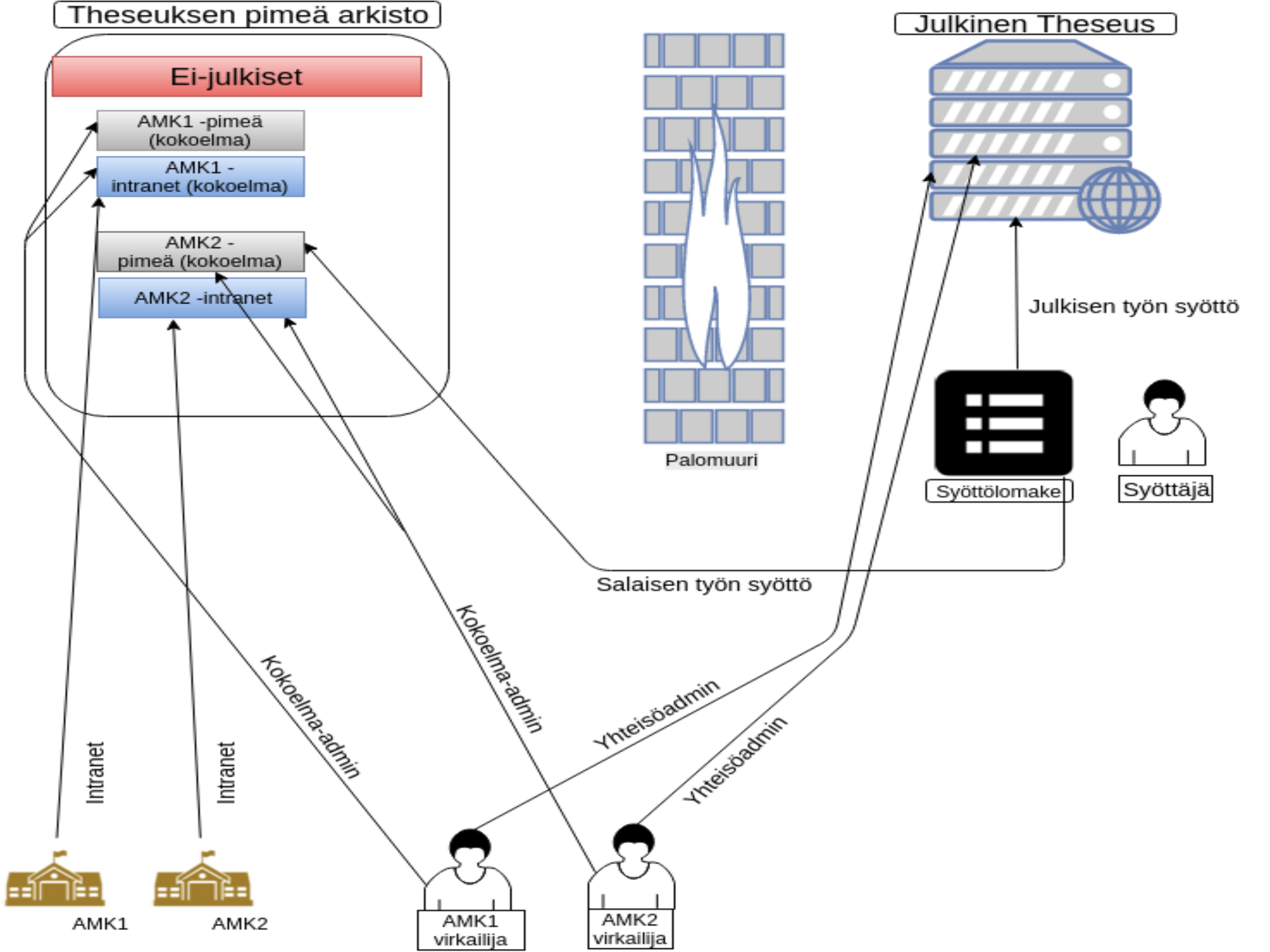
Yhteisöadmin

AMK1

AMK2

AMK1 virkailija

AMK2 virkailija



Pohdittavaa, päätettävää ja huomioita ennen toteutusta

- Julkaisuviive: Jos julkaisuviivettä ei saa olla, malli 1 vaikea toteuttaa. (Poistot ja jälkikäteen muokkaukset hankalia saada synkronoitumaan nopeasti)
- **Täysin salaisia töitä ei pimeäänkään arkistoon!**
 - Toisen kirjaston virkailija ja pääkäyttäjä pystyy lukemaan toisen amkin tietueita
 - AMK-kohtainen salaus monimutkaistaisi auktorisointien määrittelyä eikä siinä ole järkeä: Pääkäyttäjät (noin 5 kpl) pystyy kuitenkin lukemaan kaikki tietueet
 - Oikeasti salaisilla töillä ei siis ole pimeäänkään arkistoon asiaa
 - Auktorisointi mutkistaisi luovutusta PAS:iin

Pohdittavaa, päätettävää ja huomioita ennen toteutusta

- Miten lomake laaditaan siten, että suuri osa ei valitse pimeän arkiston tallennusvaihtoehtoa?
- Miten tiedonhakija saa tiedon ei julkisista julkaisuista? Jos tietue pimeässä arkistossa, metatiedot eivät haettavissa eivätkä haravoidu esim. Finnaan
- **Uusi ehdotus (malli 4) Theseus ohjausryhmältä:**
Kaikki tietueet tallennettaisiin Theseukseen, mutta ei-julkisilla olisi vain loputon embargo. Eli vain metatiedot näkyvissä näissä, pdf pääsyrajoitettu
 - + Metatiedot haravoitavissa ja haettavissa julkisesti
 - + Ei tarvittaisi erillistä pimeää arkistoa
 - Theseuksen vahvuus ollut, että tietueessa aina tiedosto mukana, ns. aito julkaisuarkisto



Kiitos!