



ensaka skattehemmans vällennämnden i Hel  
singe socknen och härad. Årlands län.  
Upprättad åren 1913-15 af  
Edvard Tiskilla.  
violantmätare.

Lägerhemmas tecken.  
R. N. 1<sup>o</sup> del. 1<sup>o</sup> Brändö  
1<sup>o</sup> Brändö Villastad.  
Införda i jordregistret 1913.  
A. Kammikainen.

Renovationens riktighet består i  
högstens vännar  
Lägerhemmas



postagare. — Näihin sarakkeisiin merkittään tiedot kuskista

Table with handwritten entries. Columns include: Kän., Sukupuoli, Koulutus, Aikamatka, Sotien tai yrk., Sotia- tai ammatt., Homot tai alust., Län. — Hälad., Kommun — Kunta, and various other administrative fields.

Table with printed headers and some handwritten entries. Headers include: I dessa kolonner införas uppgifter endast för dem som begärs sig på arbetsförordning till Sverige eller Europa. Nästa sarakkeis merkittään tiedot alustuksen nimistä, jotka liitetään työmatkalla olevien matkustajien nimistöön.



# KANSALLISARKISTO



# DALAI



**Digitaalisten aineistojen laadun ja käytettävyyden parantaminen  
tekoälyavusteisesti**

# DALAI-hanke

- 1.9.2021-31.8.2023
- Rahoittaja
  - Euroopan kehitysrahasto EAKR
  - Suomen kehitysrahasto-ohjelma, Kestävää kasvua ja työtä 2014-2020
- Partnerit
  - Kansallisarkisto (pää toteuttaja)
  - Kaakkois-Suomen ammattikorkeakoulu XAMK
  - Elinkeinoelämän keskusarkisto ELKA
  - Disec Oy

Vipuvoimaa  
EU:lta  
2014–2020



Euroopan unioni  
Euroopan aluekehitysrahasto



KANSALLISARKISTO



Kaakkois-Suomen  
ammattikorkeakoulu



SUOMEN ELINKEINOELÄMÄN  
KESKUSARKISTO

DISEC

MIKKELI

# Hankkeen kohderyhmä ja tavoitteet

- Kohderyhmä:
  - Muistiorganisaatiot ja yritykset
- Tavoitteet:
  - Kehittää tekoälyyn perustuvia menetelmiä, joilla voidaan vähentää digitoinnissa tehtävää manuaalista työtä sekä tehostaa digitointitoimintaa
  - Kehittää aineiston automaattista kuvailua eli metatietojen generointia tekoälyä hyödyntäen
  - Edistää kehitettyjen menetelmien käyttöönottoa muistiorganisaatioissa ja laajemminkin:

**Tuotantovalmiit komponentit tulevat vapaasti kaikkien hyödynnettäviksi**

# Työpaketit

- TP1: Digitoidun kuvalaadun automaattinen validointi (KA)
- TP2: Sisällöntunnistuksen kehittäminen (KA)
- TP3: Aineiston rakenteiden tunnistaminen ja automaattinen metatiedottaminen (KA)
- TP4: Yritysyhteistyö ja verkostomainen kehittäminen (Disec Oy)
- TP5: Tekoäly laajempaan käyttöön – Arkkiivi-käyttöliittymä (XAMK)

# Työpaketit

- **TP1: Digitoidun kuvalaadun automaattinen validointi (KA)**
  - Post-it lappujen tunnistus
  - Taittuneiden kulmien tunnistus
  - Tyhjien sivujen tunnistus



# Työpaketit



- **TP2: Sisällöntunnistuksen kehittäminen (KA)**

- Kirjoitustyypin tunnistaminen
- OCR laadun tutkiminen ja kehittäminen (OCR= teknologia, jonka avulla tunnistetaan koneellisesti tuotettua tekstiä)
- NLP-tekniikan testaaminen (NLP-tekniikka=luonnollisen kielen prosessointi)

- **TP3: Aineiston rakenteiden tunnistaminen ja automaattinen metatiedottaminen (KA)**

- Asiakirjakokonaisuuden tunnistaminen
- Nimientiteettien tunnistaminen
- Asiasanoitus
- Sivusegmentointi

# Työpaketit

- **TP 4: Yritysyhteistyö ja verkostomainen kehittäminen (Disec Oy)**
  - Kehittämistyöhön osallistuvan yhteistyöverkoston kokoaminen
  - Digitoitujen aineistojen kerääminen opetusaineistoksi yhteistyöverkostolta
  - Työpajojen järjestäminen ja fasilitointi
  - Tulosten jalkauttaminen verkoston muistiorganisaatioihin ja yrityksiin, komponenttien pilotointi
  - Hankkeen tuloksista kertominen konferensseissa ja tapahtumissa
- **TP5: Tekoäly laajempaan käyttöön (Xamk)**
  - Käyttöliittymän suunnittelu ja kehittäminen
  - Työpakettien 1-3 komponenttien integroiminen käyttöliittymään
  - Komponenttien käyttöön liittyvän opetusmateriaalin toteuttaminen (videot, kurssit)
  - **Arkkiivi-käyttöliittymää kehitetään tällä hetkellä ja se avataan virallisesti syksyllä 2023**



# DALAI



[www.arkkiivi.fi](http://www.arkkiivi.fi)  
<https://kansallisarkisto.fi/dalai>