

Askeleet kohti kansallista rekisteritutkijan etäkäyttöympäristöä

Marianne Johnson

Rekisteritutkimuksen tukikeskus ReTKi

www.rekisteritutkimus.fi

Lähtökohtia ja aloitteita

- Rekisteritutkimuksella on suuret mahdollisuudet
- Rekistereitä käytetään tutkimuksissa kuitenkin verrattain vähän
 - Ongelmia: käyttö lupien myöntämiseen ja aineistojen toimittamiseen kuluva aika, dokumentaation puute, monia toimijoita, aineistorajoitteet, hinta
- Joidenkin maiden tilastokeskukset ryhtyivät tarjoamaan mikroaineistoja etäkäyttöjärjestelmien kautta 2000-luvun alussa: Tanska (2001),
- Suomessa aloitteita kahdelta taholta
 - TK aloitti etäkäytön kehittämisen 2008 tutkimuslaboratorion etäkäyttöprojektina
 - ReTki esitti eri rekisterinpitäjien aineistojen yhteiskäyttöön pohjautuvaa etäkäyttöjärjestelmää (MIDRAS) tutkimusinfrastruktuurikartoituksen tiekartalle 2008
 - todettiin hankkeeksi, jolla olisi mahdollisuuksia kehittyä merkittäväksi kansalliseksi tutkimusinfrastruktuuriksi
 - CSC:n ja ReTkin valmisteluhankehakemus OKM:lle 2009

Tilastokeskuksen tutkimuslaboratorion etäkäyttöprojekti (M. Pankasalo 23.8.2010)

- Projekti alkoi 1.4.2008 ja päättyi 31.10.2009
- Tavoitteena tutkimusaineistojen etäkäyttöjärjestelmä
 - tutkijoiden tasapuolinen kohtelu
 - mikroaineistojen käytön tehostuminen
 - turvallinen ja valvottu ympäristö
 - aineistoja käytetään TK:n palvelimella
 - voidaan luovuttaa yksikötason henkilö- ja yritysaineistoja
- Mallina Tanskan ja Hollannin järjestelmät
- Toteutus aloitettiin rakentamalla projektissa testijärjestelmä tutkimuslaboratorion aineistoille
- Projektin loppuraportti on verkossa
 - <http://tilastokeskus.fi/org/tilastokeskus/vuosiohjelma.html>

Tilastokeskuksen etäkäyttöpalvelu

(A.Katainen 1.12.2010)

- Maksullinen etäkäyttöpalvelu alkoi helmikuussa 2010
- Tutkijat ottavat yhteyden työasemaltaan etäkäyttöpalvelimelle
- Palvelimella tutkijalla on käytössä Windows-työpöytä, jossa pääsee käyttämään yksikköaineistoja
- Työvälineinä STATA, SPSS, R, Acrobat reader ja tekstinkäsittelyohjelma
- Järjestelmästä ei pysty siirtämään tietoa ulos tai sisään ja tulokset saa käyttöönsä tarkistusmenettelyn jälkeen
- Tietoturva
 - Vahva tunnistaminen. Käyttäjätunnus, salasana ja GSM-tunnistautuminen
 - Palvelimet eristetty TK:n verkosta ja suojatut yhteydet rajatuista IP-osoitteista
 - Lokitietojen keräys

Tilastokeskuksen etäkäyttöpalvelun aineistot

- ”Tutkimuslaboratorion yritys ja toimipaikka-aineistot”
 - Tilinpäätösaineistopaneelit ja yritysten ulkomainen omistajuus -aineisto
 - Teollisuuden toimipaikkapaneeli
 - Hyödykkeet
 - Yritysrekisterin yritykset
 - Yritysrekisterin toimipaikat
 - T&K- ja innovaatioaineistot ja pääomakantakyselyt
 - Patenttianeistot
 - ICT-aineistot
 - Työssäkäyntitilaston yritystason henkilöstöominaisuudet
 - Työssäkäyntitilaston toimipaikkatason henkilöstöominaisuudet
 - FLEED-otos
 - Yritystukiaineisto
- Laajennettu koskemaan muita Tilastokeskuksen aineistoluovutuksia
 - Henkilöaineistoja ja erilaisia yhdistelmäaineistoja

Hanke: MIDRAS- rekisteritietojen etäkäyttö

- **Aikataulu: 4/2009 – 4/2011**
- **Toteuttajat:**
 - CSC- Tieteen tietotekniikan keskus
 - ReTki – Rekisteritutkimuksen tukikeskus
- **Rahoitus:**
 - OKM
- **MIDRAS selvitys- ja valmisteluhanke**
 - Rekisterinpitäjien sitouttaminen, hallinnollisten ja lainsäädännöllisten kysymysten selvittäminen
 - Etäkäyttöjärjestelmän tietotekniset ratkaisut: IT-arkkitehtuuri, prosessit, käyttötapaukset
 - Pilottijärjestelmä (toteutus, testaus)
 - Loppuraportti:
 - MIDRAS: Viranomaisten luvanvaraiset aineistot tehokkaasti ja turvallisesti tutkimuskäyttöön
 - <http://www.csc.fi/sivut/e-infra/midras/>

Midras -järjestelmän toimintamalli

- Etäkäyttöjärjestelmä tarjoaa palveluita kirjautuneille, käyttöluvan omaaville käyttäjille
- Aineistojen omistajuus säilyy rekisterinpitäjillä
- Rekisterinpitäjät päättävät siitä, millä tavoilla ja aineistoilla he osallistuvat etäkäyttöjärjestelmään.
- Palveluna on tietojen siirto- ja käyttörajapinta MIDRAS-ylläpito – organisaation ja rekisterinpitäjän välillä.
- Kutakin aineistoa käsitellään sitä koskevien lakien mukaisesti.
- Tietoturva- ja suoja ovat keskeisessä asemassa MIDRAS-ympäristön suunnittelussa.



MIDRAS-etäkäyttöjärjestelmä

Aineistojen käyttö lupaa ja tunnistautumista edellyttäviä palveluja:

- Etätyöpöytä aineistojen käsittelyyn (ohjelmistot ja työkalut)
- Projekteittain eristetyt aineistot kuvailutietoineen
- Tiedostojen luovutus- ja lisäyspalvelu

Tunnistautumista edellyttäviä palveluja:

- Keskitetty sähköinen lupahakemuspalvelu

Julkisia palveluja:

- Metatietokatalogi
- Tutkimuksen ja opetuksen tukipalvelut

Aineistojen ja metatietojen rajapintapalvelut, pseudonymisointi

Käyttöoikeuksien hallintapalvelut

Aineistojen toimitus verkon kautta



Sähköinen hakemusten käsittely

TURE-ryhmä:

Tilastokeskuksen ja Midras- projektin etäkäyttäjärjestelmien yhteiskehittämistä valmisteleva työryhmä

- Valtiovarainministeriö asettama 31.5.2010 -30.6.2011 (lisäaika 30.9.2011)
- Tavoite:
 - Suomessa tarjotaan hallinnollisten aineistojen sekä tilastorekisterien ja – aineistojen tietoja tutkijoille yhtenäisen etäkäyttäjärjestelmän kautta, joka on osa kansallista tutkimuksen e-infrastruktuuria.
- Työryhmän tehtävänä oli:
 - selvittää Tilastokeskuksen ja Midras-projektin etäkäyttöhankkeiden yhteiskehittämis-mahdollisuudet,
 - selvittää ja tehdä ehdotus etäkäyttäjärjestelmien hallintomalliksi sekä
 - selvittää ja tehdä ehdotus järjestelmien toteuttamisen ja ylläpidon resursoinniksi ja rahoitukseksi.
- Loppuraportti julkaistu valtioneuvoston hankerekisterissä (http://www.hare.vn.fi/upload/Asiakirjat/16576/176311_Ture_loppuraportti_V99.pdf)
- Loppuraportti lausunnolla 21.12.2011 asti

TURE-ryhmä

- Työryhmän puheenjohtajana toimi johtaja Pirjo-Leena Forsström CSC
- Työryhmän jäsenet:
 - neuvotteleva virkamies Arja Terho VM,
 - erikoistutkija Markku Suvanen OKM,
 - neuvotteleva virkamies Sari Kauppinen (erityisasiantuntija Ilari Keso) STM,
 - tilastopäällikkö Jussi Heino TK,
 - päällikkö Irma-Leena Notkola ReTki,
 - tietohallintojohtaja Jaason Haapakoski THL,
 - erikoistutkija Jouko Verho Kela,
 - professori Markus Jäntti Tukholman yliopistosta (tutkimusjohtaja Mika Maliranta ETLA).
- Sihteerit:
 - erikoissuunnittelija Marianne Johnson ReTki
 - yliaktuaari Markku Pankasalo TK.

Etäkäyttöjärjestelmän palvelut

1. Metadatakatalogi

- Eri rekisteriorganisaatioiden aineistot
- Tuotettu yhteisellä metatietostandardilla
- Hakumahdollisuus
- Myös vanhat aineistot

2. Käyttölupien haku- ja hallintapalvelut

- Yhteinen sähköinen 'lupalomake'
- Ohjaava
- Yhteys metadatakatalogiin

3. Aineistojen ja metatietojen rajapintapalvelut

- Vastaanottopalvelu etäkäyttö-järjestelmässä
- Esimerkkityökaluja:
 - Aineistojen pseudonymisointiin
 - Aineistojen lähettämiseen
 - Aineistojen muodostamiseen käyttölupahakemuksen mukaiseksi

Etäkäyttöjärjestelmän palvelut (jatk)

4. Etäpöytäympäristöpalvelut

- Tutkimus, raportointi ja analyysiohjelmanpalvelu
- Käyttäjätunnuksella pääsyä
- Omalla virtuaalipalvelimella oma tutkimusaineisto
- Ohjelmat: SAS, SPSS, STATA, R

5. Omien tiedostojen lisäys- ja luovutuspalvelu tarkastuksen kautta

- Keino siirtää tietoja ja tiedostoja etätyöpöydän ja muun maailman välillä
- Mahdollinen viranomaisen tarkastuspiste luovutuspalvelussa

6. Tutkimuksen ja opetuksen tukipalvelu

- Help desk
- Keskustelu- ja tukifoorumi
 - Hyväksi havaitut käytännöt, metatietojen parannusehdotukset, aineistojen korjauspyynnöt, aineistojen käsittelyssä käytetyt menetelmät (työkalupakki)

Työryhmän päätelmät

- Vaikka tutkimuksen edistäminen on opetus- ja kulttuuriministeriön vastuulla, rekisteritutkijan etäkäyttöympäristön kehittäminen vaatii useiden hallinnonalojen ja toimijoiden yhteistyötä.
- Rekisteritutkijan etäkäyttöympäristön pysyvään toimintaan tarvitaan hallinnollinen toimija (viranomainen), joka vastaa palvelun kehittämisestä, toiminnasta ja sopimuksista ja toimii tietoteknisten palvelujen tilaajana.
- Työryhmä ei ole päässyt yhtenäiseen näkemykseen hallinnollisesta toimijasta ja ehdottaa tässä vaiheessa väliaikaisena ratkaisuna, että työtä ryhdytään edistämään komponenteittain, tavoitteena MIDRAS-mallin mukainen rekisteritutkimuksen etätyöpöytäympäristön kehittäminen.
- Työryhmä esittää väliaikaiselle ratkaisulle aikataulun vuoteen 2013, jolloin pyritään pääsemään pysyvään toimintamalliin. Työryhmä ehdottaa että opetus- ja kulttuuriministeriö kokoaa yhteistoiminnalle ohjausryhmärakenteen.

Kehittämiskomponentit

Työryhmä esittää seuraavia tehtäviä ja vastuutahoja vuosille 2011-2013:

Luodaan sähköinen lupahakupalvelu tutkijoille

Aikataulu: 2012-2013, valmis 1H2013.

Tulos: Sähköinen lupapalvelu, jonka kautta tutkijat voivat hakea käyttö lupia eri viranomaisilta ja joka on liitettävissä metatietokatalogiin.

Kehitysvastuu: ReTki

Kehitetään rekisteritutkijoiden etätyöpöytäympäristöä.

Aikataulu: syksy 2011-2012, valmis 4Q2012.

Tulos: Tutkijaa mahdollisimman laajasti palveleva etätyöpöytäkonsepti työvuopohjaisine palveluineen.

Kehitysvastuu: Tilastokeskus

Kehitetään metadatakatalogipalvelua CSC:n tutkimusaineistojen metatietokatalogityön kautta.

Aikataulu: käytettävissä 2H2012.

Tulos: Metadatakatalogipalvelu metatietojen jakamiseen.

Kehitysvastuu: CSC

Kehittämiskomponentit (jatk.)

Perustetaan web-pohjainen käyttäjäfoorumi,

jonka kautta rekisteritutkijat voivat osallistua etätyöympäristön suunnitteluun ja kehittämiseen. Myös tutkijapaneeleita järjestetään.

Koordinointivastuu: ReTki

Edistetään aineistojen kuvailua

a) laatimalla JHS-työryhmän kautta rekisteriaineistojen metatietojen sisällön ja säilytysmuotojen standardisointi) ja b) tukemalla aineistojen kuvailutyötä metatietostandardin mukaisesti standardin valmistuttua.

Kootaan hanketta ohjaava, tutkimuksen näkökulmaa edistävä ohjausrakenne.

Perustetaan yhteistyöfoorumi,

jossa best practises -toimintaa edistetään rekisteriviranomaisten ja palveluorganisaatioiden ja ylläpitäjien kesken.

Osallistutaan 2013 tutkimusinfrastruktuurikartoitukseen.

Jos hankkeelle saadaan rahoitusta, laaditaan pysyvä hallinnointimalli palveluille.

Edistetään rekisterien muokkaamista ns. valmisaineistoiksi