

# FiLRC: FIN-CLARIN-johtoryhmän kokous

Aika: pe 4.3.2011 klo 13-16.00

Paikka: Is 5, Helsingin yliopisto, Metsätalo, Unioninkatu 40

## PÖYTÄKIRJA

Kokous alkoi klo 13.05

### Osanottajat:

HY: Kimmo Koskenniemi (pj), Mietta Lennes, Krister Lindén, Jussi Piitulainen, Hanna Westerlund (siht.)

CSC: Tero Aalto, Antti Pursula

JY: Ari Huhta

Kotus: Antti Leino, Pirkko Nuolijärvi (klo 14 alkaen), Toni Suutari

OY: Marketta Harju-Autti

TaY: Mikhail Mikhailov

### 1. Yleistä - ilmoitusasioita

Muutoksia FIN-CLARIN johtoryhmään tulossa: Oulun yliopiston uudet edustajat ovat varsinainen jäsen Jarmo Harri Jantunen ja varajäsen Harri Mantila. Tervetuloa! Muutos tulee voimaan, kun nykyinen edustaja Marketta Harju-Autti jää eläkkeelle 1.8.2011. Kiitokset Marketalle!

Johtoryhmätyöskentelyyn toivottiin tervetulleeksi uusi CSC:n edustaja Antti Pursula.

### 2. CLARIN ERIC

Perustusvaiheessa oleva CLARIN ERIC koostuu kolmesta tasosta: Hallinnollisella tasolla hanketta hoitaa keskushallinto toimistoineen Haagissa. Toinen taso sisältää teknisen infratruktuurin tason, jota Suomessa edustaa CSC. Sisällöntuotannosta vastaavat kansalliset konsortiot ovat kolmas taso, jota Suomessa edustaa FIN-CLARIN. Käytännössä Suomessa kaksi viimeksimainittua tasoa toimivat molemmat FIN-CLARIN -konsortion sisällä.

### 3. FIN-CLARINin vuosikertomukset

Vuosikertomusten 2009 ja 2010 luonnokset ovat näkyvissä FIN-CLARIN hallinnon sivulla ja niihin toivotaan kommentteja.

Käynnissä olevat hankkeet: HFST, [FinnWordNet](#), [FinnTreeBank](#) (HY yhdessä Kotuksen kanssa), Kielipankin päivitys (koko FIN-CLARIN).

### 4. FIN-CLARIN CONTENT

Suomen Akatemian hankerahoitus ajalle 4.1.2011-30.4.2015 suunnataan pääosin olemassa olevien aineistojen saamiseksi FIN-CLARINin piiriin ja Kielipankin kautta laajasti tarjolle. Johtoryhmän tehtävänä on päättää, mitä aineistoja hankerahoituksella päivitetään ja missä järjestyksessä hankerahoitus käytetään. Tarvitaan tarvekartoitus.

Ensimmäinen hanke, jonka osalta neuvottelut on aloitettu, on Kotuksen murreaineisto (puheaineisto), josta Toni Suutari kertoi tarkemmin. Aineisto sisältää puheäänitteitä sekä murrekirjoista skannattuja litteraatioita, joita on jouduttu käsin korjaamaan.

Tarkastustyö on lähes valmis. Lisäksi suuri osa aineistosta on yleiskielistetty tutkimuskäytön ja hakujen helpottamiseksi. Myös äänitteiden ja litteroinnin kohdistaminen on suunnitteilla. Nyt selvitetään tekijöitä, jotka vielä kohdistamistyön lisäksi olisivat FIN-CLARINin näkökulmasta katsottuna syytä ottaa huomioon ja ratkaistaan tarvittaessa esiin tulevat tekniset kysymykset. Selvitettäviin asioihin kuuluvat mm. hakuliittymään liittyvät kysymykset. Lisäksi tarvitaan käyttäjien ohjeistus.

Kimmo Koskenniemi käy lisäkeskusteluja Kopioston edustajien kanssa mahdollisuudesta saada laajoja tekstiaineistoja FIN-CLARINin palveluiden piiriin. Sen lisäksi, että aineistoja, joiden tekijänoikeus on rauennut, voidaan saada mukaan aineistotarjontaan, keskustellaan myös tekijänoikeuden alaisista aineistoista ja siitä, millä ehdoilla niitä voisi saada käyttöön. FIN-CLARIN voisi ehkä saada Kopiostolta sopimuslisenssin tietyin ehdoin, esim. vain luotettavasti tunnistettujen käyttäjien saataville (esim. HAKA-tunnistautumisen kautta).

### 5. META-NET ja META-NORD

Krister Lindén kertoi META-NETistä [www.meta-net.eu](#). Se on yhteenliittymä, jonka alla on kahdeksan EU:n seitsemännen puiteohjelman projektia ja josita META-NORD on yksi. Tavoitteena on toteuttaa CLARINissa jo aloitettu työ kieliresurssien saamiseksi tarjolle koko Euroopassa ja kielentutkimuksen tärkeyden osoittamiseksi. META-NET tarjoaa mm. metadatan varten portaalin META-SHARE, jonka kautta käyttäjät tulevat saamaan tietoa resurssien olemassaolosta ja saatavuudesta. Siinä, kun CLARIN keskittyi kielitieteellisten resurssien kartoittamiseen ja tekijänoikeuksien ja standardien kartoittamiseen tutkimusta ja opetusta varten, sisältyy META-NET -hankkeeseen myös kaupallisten toimijoiden tarpeiden kartoittaminen ja resurssien tarjoaminen kaupallisten kieliteknologisten sovellusten kehittämiseen.

META-NORD [www.meta-nord.eu](#) puolestaan toimii Baltian maiden ja Pohjoismaiden voimin. Sen tavoitteena on pienten kielten aseman kartoittaminen sekä aineistojen (metadatan) tuominen käyttäjien näkyville ja kielentutkimuksen tärkeydestä tiedottaminen. FIN-CLARINin rooli tiedon tuottajana, keskusteluun osallistujana ja resurssien tarjoajana on merkittävä edellytys projektin onnistumiselle. Kaikki META-NORDin tavoitteet ovat FIN-CLARINin tavoitteiden kanssa yhteneväisiä. Kevään kuluessa tullaan kartoittamaan suomen kielen tilaan liittyviä kysymyksiä, joissa Kotuksen ja CSC:n toimittamat raportit ja strategiat ovat keskeisiä, mutta lisäksi tarvitaan kaikkien asiantuntemusta suomen kielen tilasta kertovan raportin koostamisessa. Aineistojen kartoittaminen on jo FIN-CLARINin puitteissa kertaalleen tehty, joten nyt riittää tiedon päivittäminen ja uusien resurssien lisääminen. META-NORD ja CLARIN tarjoavat aineistojen sijoittamisen ja sijoitussopimusten päivittämisen tukea tarvittaessa. Resurssin tuominen tarjolle META-NORDin kautta ei edellytä sen siirtämistä, sen käyttöoikeuksien muuttamista, vaan riittää, että aineistosta on tehty metadata, joka sisältää kaikki potentiaalisen käyttäjän tarvitsemat tiedot ja että metadata asetetaan julkiseksi esim. CSC:n palvelimen tai

META-SHAREn kautta. Metadata voi myös olla tarjolla useampaakin kautta, mutta yhteisesti todettiin, että on järkevää huolehtia siitä, ettei metadata-tietueita ole useampia eri paikoissa, jolloin päivittämisestä saattaa tulla hankalaa. Hannan ppt-tiedosto on alla liitteenä.

Language Report -tietojen keräämistä varten tälle samaiselle FIN-CLARIN hallinnon sivulle on tehty linkki, jonka taakse voivat kaikki kirjata tietojensa kysytyistä asioista. Kaikki tieto on tervetullutta! Lopulliseen raporttiin haluttaisiin myös useita kirjoittajatahoja ja kommentaattoreita.

## 6. Kielipankki

Todettiin hyväksi Kimmon esityslistassa hahmottelema ja Antti Pursulan tukema malli työnjaosta FIN-CLARIN -konsortion sisällä. CSC tarvitsee FIN-CLARINilta selkeän viestin siitä, mitä palveluja tarvitaan ja millaisia resursseja Kielipankkiin halutaan. Työnjako täytyy johtoryhmän ohjauksella täsmentää ja priorisointi dokumentoida. Tarvitaan tarvekartoitus.

AAI on CSC:n kehittämä autentikointi -ja auktorisointijärjestelmä, joka tulee korvaamaan CSC:n nykyisen käyttöluvan hakumenettelyn. AAI on tarjolla HAKA-autentikoinnin takaa Tutkijan käyttöliittymästä (SUI = Scientist's User Interface). Näin kaikki HAKA-tunnuksen omistajat voivat kirjautua sisään Tutkijan käyttöliittymään, josta pääsee kuvakkeen kautta hakemaan haluamiaan käyttöluvia Kielipankin aineistoihin. Järjestelmä toteuttaa kaiken hakumenettelyyn liittyvän sähköisesti ja kun tunnus on luotu, lähettää CSC sen hakijalle postin kautta. Muita vaihtoehtoja toimittaa lupa tutkitaan. CSC:n käyttöluvan haltijat pääsevät myös sen avulla hakemaan samaisen AAI-sovelluksen avulla lisää käyttöluvia.

## 7. Mahdollinen yhteinen metadata-työpaja ja/tai paikalliset työpajat

Kotus järjestää FIN-CLARIN metadata-työpajan yhdessä HY:n kanssa. Kotuksen yhteyshenkilö on Toni Suutari ja HY:n puolesta asiaa hoitavat toistaiseksi Hanna Westerlund ja Jussi Piitulainen. Kaikkia pyydetään haarukoimaan sopivaa ajankohtaa ja raportoimaan epäsovivista. Tarkoitusta varten avataan Doodle-taulukko. Selvitetään olisiko mahdollista järjestää jo toukokuussa tai kesäkuun alkupuolella.

## 8. Muut asiat

Todettiin vielä, että FIN-CLARINista tarvittaisiin jonkinlainen esite tai poster. Esim. A4-kokoinen tiivis info olisi hyvä lisä johtoryhmän jäsenten käyttöön, kun FIN-CLARINista tiedotetaan paikallisesti tai esim. konferensseihin ja muihin tapahtumiin voisi vielä tietoa mukanaan.

## 9. Seuraava kokous

Metadata-työpajan yhteydessä tavataan henkilökohtaisesti ja sitä ennen voidaan pitää verkkokokous tarpeen mukaan.

## Tehtävälista

### Mitä / kuka / milloin

- 1) Tarvekartoitus FIN-CLARIN CONTENT / kuka?/ miten? /milloin?
- 2) Tarvekartoitus CSC / kuka? /miten?/ milloin?
- 3) Suomen kielen tila -raportti (language report) ja resurssikartoitus / Hanna & tiimi Hissu Hyvärinen ja Pinja Pennala ottavat yhteyttä resurssilistan päivitystä varten /maaliskuu
- 4) Metadata-työpaja, ajankohdan selvittäminen / Hanna ja Toni / mahdollisimman pian.
- 5) Mitä konferensseja yms. tapahtumia on toukokuussa ja kesäkuun alkupuolella ? / kaikki / miettikää valmiiksi!
- 6) Esite tai muu materiaali. Löytyisikö vapaaehtoisia? Posterit yms. 1.4. jälkeen mahdollisia.

[MiettaLennes](#) pitää esitteiden ja postereiden taittamisesta ja ilmoittautuu vapaaehtoiseksi.

Kokous päättyi klo 16.02.

-- [HannaWesterlund](#) - 2011-03-10