

Hallinto: FIN-CLARIN -johtoryhmän kokous

Aika: ma 20.2.2012

Skype-neuvottelu

PÖYTÄKIRJA

Osallistujat:

Tero Aalto, CSC

Ari Huhta, JY

Kaisa Häkkinen, TY

Nobufumi Inaba, TY

Merja Koskela, VY

Kimmo Koskenniemi, HY/FIN-CLARIN, puheenjohtaja

Mikko Laitinen, JY

Mietta Lennes, HY/FIN-CLARIN, sihteeri

Krister Lindén, HY/FIN-CLARIN

Mikhail Mikhailov, TaY

Ville Savolainen, CSC

Toni Suutari, Kotus

Kokous alkoi klo 14:01.

1. CLARIN ERIC ja FIN-CLARIN-konsortiosopimus

[Konsortiosopimus](#) etenee.

Kierros lakimiehillä jatkuu. Pyritään saamaan tämä vaihe mahdollisimman nopeasti päätökseen. Sopimusneuvotteluissa on ollut vielä joitakin avoimia kohtia, joihin yliopistojen lakimiehet ovat halunneet puuttua. Jyväskylän yliopiston lakimiehiltä odotetaan vielä kommentteja. On hyvä, että tarkennukset tulevat tässä vaiheessa, jotta sopimukset saadaan ajoissa kuntoon ennen rehtoreiden allekirjoituksia.

Huhtikuun lopussa tai toukokuun alussa on CLARIN ERICin perustamiskokous. FIN-CLARIN ei ehtine perustajajäseneksi, mutta sopimukset pitää kuitenkin saada valmiiksi perustamiskokoukseen mennessä.

2. Aineistojen priorisointi

Taustatietona oli edellisessä kokouksessa käsitelty [aineistonhankintakehys](#).

Liite: [Ehdotetut aineistot ja niiden vaatimat työmäärät](#):

Jyväskylässä meneillään kaksi hanketta, TOPLING ja DIALUKI, joissa tuotetaan kieliaineistoja

Eurooppalaisen viitekehityksen taitotasot käytössä; mukana aika paljon tausta-aineistoa, kyselyjä yms.

digitalisointi meneillään (puhtaaksikirjoitus), sanastoanalyysia varten pitää korjata oikeinkirjoituksen virheitä

alustavasti digitoidaan hankkeita varten otoksena noin viidesosa koko materiaalista; noin 16000 tekstiä digitalisoimatta

yksittäisten tekstien pituus/sanamäärä vaihtelee lyhyistä viesteistä pitempiin kirjoitelmiin

Tampere

olemassa hyvä alusta, jossa aineistot turvassa

sopimusasioita pitäisi kuitenkin jatkon varalta varmistella

Turku: Lauseopin arkiston murrekorpus

tehtävä vielä hakusanointu, harjoittelijoiden puheenvuorojen kohdistaminen

kyseessä ison aineiston täydentäminen

kaikki XML-muodossa, äänitteet muunnettava WAV-muotoon

voitaisiin saada yhtenäisen käyttöliittymän alle (LAT-alusta)

Kotus

Vanhan kirjasuomen korpus (2010-)

tekstintunnistuksen oikoluku ja rakenteistuksen hienosäätö puuttuu

kuvat voitava liittää mukaan

ongelmana kahden erillisen, korjaamattoman ja korjatun version yhteensulauttaminen

täydennettynä 7-8 milj. sanaa

Luvanvaraiset aineistot Kielipankkiin

suuri määrä pieniä aineistoja, joita voidaan antaa vain tutkijoiden käyttöön, mutta Kotuksen palvelimella käyttöoikeuksien

myöntäminen hankalaa. Tässä Kielipankki voi auttaa.

Johtoryhmän tiedoksi tuotiin myös muita aineistoja, joista on neuvoteltu:

Myös Suomalaisen kirjallisuuden seuran (SKS) kanssa on keskusteltu aineistoista, esim. kansanrunoutta (arvio vajaat 10 miljoonaa sanaa)

Kopioston kanssa neuvoteltu sanomalehti- ja aikakauslehtiaineistojen lisensoinnista, lähinnä 1900-luvun aineistoa (noin 1 miljardi sanaa)

Keskusteltiin julkaistaviksi ehdotetuista aineistoista, niiden priorisoinnista ja rahoituksesta.

Käytettävissä noin 80000 €, joka on käytettävä (maksettava FIN-CLARINin tililtä suorittavalle taholle) vuoden 2012 aikana.

Jyväskylän aineisto uudentyypistä verrattuna Kielipankin nykyisiin kokoelmiin.

Tampereen aineisto: laadukkaat rinnakkaiskorpuksukset olisivat aineistotyyppinä tärkeitä.

Myös Vanha kirjasuomi olisi Kielipankissa uusi aluevaltaus, ja tarvittaisiin jatkoa aiemmin saatavilla olleelle historialliselle

sanomalehtiaineistolle.

Kopioston neuvottelut: aineistoista maksettava korvaus on 5000 € per vuosi. Aineistoa voi käyttää vain konkordansseina, ei kokoteksteinä. Tämä kuitenkin riittää moniin tarkoituksiin, esimerkiksi sanaston tutkijalle.

SKS:n aineistojen kustannuksista ei vielä ole tietoa, keskustelut sen osalta jatkuvat.

Harkitaan mahdollista kuluvarausta sopimuskysymysten selvittämiseen.

Kaikkiin ehdotettuihin aineistoihin suhtaudutaan periaatteessa myönteisesti.

Tarkennetaan ja yhdenmukaistetaan ehdotettujen aineistojen suunnitelmat ja eritellään niihin ALV:n osuus. Vasta tämän jälkeen voidaan päättää rahanjaosta.

Täsmennettyjen hakemusten takaraja on noin kuukauden kuluttua, samoihin aikoihin kuin konsortiosopimuksen.

Hakemukset lähetetään Mietalle, joka auttaa myös hakemusten tekemisessä. Myös Kimmo ja Krister auttavat tarvittaessa.

Tässä ehdotettuja aineistoja voidaan ryhtyä edistämään heti, kun suunnitelma on riittävän täsmällinen.

3. Muut asiat

Ei muita asioita.

4. Seuraava kokous

Seuraava kokous noin kuukauden kuluttua (sovitaan [Doodlella](#), ehdotetut ajat ma 19.3. alkaen).

Kokous päättyi klo 15:27.

-- [MiettaLennes](#) - 2012-03-05

 **Kommentit**

Kommentin jättäminen vaatii [rekisteröitymisen](#) ja kirjautumisen.